DOCUMENT RESUME

ED 353 978 IR 016 014

AUTHOR	Law, Margaret H.; Rosen, Bruce K.
TITLE	Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records.
INSTITUTION	National Inst. of Standards and Technology (DOC), Washington, DC. National Computer Systems Lab.
SPONS AGENCY	National Archives and Records Administration, Washington, DC.
PUB DATE	Mar 89
NOTE	178p.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Archives; Computer Networks; *Computer Storage Devices; *Databases; *Government Publications; Information Management; Information Retrieval; Information Storage; Microcomputers; Optical Disks; *Policy Formation; *Preservation; *Records Management; Standards
IDENTIFIERS	*National Archives and Records Administration

ABSTRACT

The objectives of this project were to assist the National Archives and Records Administration (NARA) in developing a policy for the representation, transfer, access, and preservation of electronic records of permanent value. A lack of attention to the management, retention, and selective preservation of these machine-readable records will result in their loss. Even if the information is physically present on a physical storage device, without proper indexing or descriptive information, it will become unusable. A major portion of this report is devoted to discussing the internal electronic information management requirements that NARA must satisfy to preserve its electronic information holdings. It is noted that this process will be complicated by two factors: much of the descriptive information that NARA must manage must be provided by the originating agencies, and constant changes in computer and storage technology will require constant monitoring and readiness to change methods of managing electronic record holdings. Various standards that now exist, or are expected to exist, are recommended for use by NARA in seeking solutions to the problems of managing electronic records. A prototype software system for the exchange of documents produced under different word processor systems is presented to illustrate the application of these standards. Four appendices provide additional detailed information: (1) Document Interchange Standards: Description and Status of Major Document and Graphics Standards (Judi Moline, 1988); (2) Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status (Margaret H. Law with assistance from Leonard Gallagher and Tim Boland); (3) Recommendations for Document Transfer Standards and Their Integration into National Archives Policy (Judi Moline); and (4) Recommendations for Database and Data Dictionary Standards and Their Integration into National Archives Policy (Wilma M. Osborne, Bruce Rosen, and Leonard Gallagher). (DB)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED353978

Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records

Prepared by the

National Computer Systems Laboratory
National Institute of Standards and Technology

For the

National Archives and Records Administration

By Margaret H. Law
and Bruce K. Rosen

March, 1989

IR016014

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Judi Moline

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records

TABLE OF CONTENTS

Executive Summary 1

1 Overview of the Task 2

- 1.1 Purpose 2
- 1.2 Impact of Framework on NARA mission 2
- 1.3 Task Statement 3

2 Introduction to Software Standards and Their Development 3

- 2.1 De Facto and Consensus "Standards" 3
- 2.2 Standards Approved by Accredited Standards Organizations 4
  - 2.2.1 Accredited Standards Organizations . . . 4
  - 2.2.2 Approved Software Standards 5
- 2.3 The Importance of OSI in Current Standards . . 6
  - 2.3.1 OSI Standardization 6
  - 2.3.2 Importance of OSI for Interoperability . 6
  - 2.3.3 The Role of Protocols in OSI 7

3 Conceptual Framework for the Exchange of Electronic Records 8

- 3.1 Record Transfer 8
  - 3.1.1 Record Scheduling, Transfer, Accession, and Retrieval 9
  - 3.1.2 Dissemination of Records to Non-NARA Users 11
  - 3.1.3 Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination 14
- 3.2 Descriptive Information Capture and Interchange . 16
  - 3.2.1 Transfer of Scheduling and Descriptive Information from Creating Agencies . . . 16
  - 3.2.1.1 An Example Problem: Database Transfer . 16
  - 3.2.1.2 An Example Solution: Database Descriptive Information 18
  - 3.2.1.3 NARA Policy Needed for a Coordinated NARA-wide Data Administration Function . 18
  - 3.2.1.4 NARA Policy Needed for Accessible Descriptive Information Management . . 19
  - 3.2.2 Internal Interchange of Scheduling and Descriptive Information within NARA . . 19

3.2.2.1 Internal Information Availability . . . 20
3.2.2.2 Data Dictionary Administration Support . 22
3.2.2.3 Limits to Internal Information Creation and Modification 22
3.2.2.4 NARA Policy Needed for Data Administration Support 23
3.2.3 Queries for Record Access Via Data Administration System 23
3.2.4 Descriptive Information Interchange Framework 24
3.2.4.1 NARA Policy Needed for Capture of Scheduling Information Directly from the Creating Agencies 27
3.2.4.2 NARA Policy Needed for Cross-Referencing Information 27
3.2.4.3 NARA Policy Needed for Electronic Record Storage Information 27
3.3 Framework for Record Accession/Transfer and Descriptive Information Interchange 28

4 Creating Agency's Framework View 30

4.1 NARA Policy Needed for Shared Responsibility for Electronic Interchange 30
4.2 NARA Policy Needed for Record and Descriptive Information Format and Media 31
4.2.1 Guidelines Needed for Descriptive Information Received from Creating Agencies 31
4.3 Responsibilities of the Creating Agencies . . . 32
4.3.1 Creating Agency Access to Accessioned Records 32
4.3.2 Loss of Accessioned Records through Misidentification 33
4.3.3 Role of Data Dictionary System in Record Schedule Checking 33
4.3.4 Creating Agency Responsibilities for Accurate Information Interchange 34
4.3.4.1 Creating Agency Participation in NARA's Information Interchange System 34
4.3.4.2 Provision of Adequate Record Schedule and Descriptive Record Information . . 34
4.3.4.3 Provision of Records in Acceptable Electronic Record Formats 34
4.3.4.4 Provision of Records in Acceptable Electronic Media 35

5 NARA's Framework View 35

5.1 Accessioning Policy for Formatted Records and Associated Descriptive Information 35

5.2 Preservation of Stored Records 36
5.3 Media Assignment 37
5.4 Format and Media Conversion after Accession/Transfer 37

6 Future User Framework View 38

6.1 Record Access and Retrieval Information . . . 38
6.2 Record Access Format and Media 39
6.3 Record Preservation 39

7 Data Administration Framework View 40

7.1 Centralized Coordination of Data Administration . 40
7.2 Methodology and Automated Support for Data Administration 40
7.3 Descriptive Information Interchange 41

8 Document Standards 41

9 Form Standards 42

10 Graphics Standards 42

11 Database Standards 43

12 Library Interchange Standard 44

13 Generic Interchange Standards 44

13.1 Abstract Syntax Notation One (ASN.1) 44
13.2 Data Descriptive File (DDF) 45

14 Additional Recommendation Information 45

14.1 Responsibility of NARA to Define Interchange Policy 45
14.2 Responsibility of NARA to Define Data Administration Policy 47
    14.2.1 Assessment of NARA Data Administration Position 47
    14.2.2 Need for Coordination of Data Administration Activities 47
    14.2.3 Need for Data Administration Support for Record Storage and Access 47
    14.2.4 Need for Descriptive Information Automation 48
14.3 Responsibility of NARA to Define Policy for Systems Integration 48

LIST OF FIGURES

Figure 1 Record Scheduling, Transfer, Accession, and Retrieval 10

Figure 2 Dissemination of Records to Non-NARA Users . . 12

Figure 3 Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination 15

Figure 4 Transfer of Scheduling and Descriptive Information to NARA from Creating Agencies . . 17

Figure 5 Internal Interchange of Scheduling and Descriptive Information within NARA 21

Figure 6 Queries for Record Access Via NARA Data Administration System 25

Figure 7 Descriptive Information Interchange Framework . 26

Figure 8 Record Transfer/Accession and Descriptive Information Interchange Framework 29

ATTACHMENTS

Attachment A	Document Interchange Standards: Description and Status of Major Document and Graphics Standards, NISTIR 88-3851, Sept 1988
Attachment B	Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status
Attachment C	Recommendations for Document Transfer Standards and Their Integration into National Archives Policy
Attachment D	Recommendations for Database and Data Dictionary Standards and Their Integration into National Archives Policy

EXECUTIVE SUMMARY

This report is the result of a project conducted by the National Computer Systems Laboratory of the National Institute of Standards and Technology (NIST) (Formerly NBS) at the request of the National Archives and Records Administration (NARA). Computers are being utilized to create, process and store ever increasing amounts of government documents and database information. As time goes on, information and documents that appear on paper now will exist only in electronic form on storage devices such as magnetic disk, optical disk, etc. A lack of attention to the management, retention, and selective preservation of these machine-readable records will result in their loss, possibly for all time. If this is allowed to happen, then the U.S. Government will in essence "lose its memory." This problem is further complicated by the fact that even if the information is physically present on a physical storage device, without proper indexing or descriptive information, the information will become unusable, and thus the effect is the same as if it was never saved. It is this descriptive information that is the key that unlocks electronic records for retrieval and use. Accordingly, a major portion of this report discusses the internal electronic information management requirements that NARA must satisfy if it is to prevent the loss of its electronic information holdings. This report also points out that much of the descriptive information that NARA must manage, must be provided to NARA from the originating agencies in a manner consistent with the policies that must be established by NARA. Also discussed is the additional complication that, constant changes in computer and storage technology will require that NARA constantly monitor, and be prepared to change its methods of storing, maintaining, and retrieving electronic record holdings. Finally, throughout this report various standards that now exist, or are expected to exist, are recommended for use by NARA in seeking solutions to the problems of receiving, storing, maintaining and retrieving electronic records. As an example of the application of these standards, a prototype software system for the exchange of documents produced under different word processor systems, was also developed as a part of this project.

1 Overview of the Task

This section provides an overview of the task performed by the National Institute of Standards and Technology (NIST) for the National Archives and Record Administration (NARA), to develop a framework for the interchange and preservation of NARA's electronic records.

1.1 Purpose

The objective of this framework is to assist the National Archives and Records Administration (NARA) in developing a National Archives Policy for the representation, transfer, access, and preservation of electronic records of permanent value. The framework identifies the national and international computing software standards to be considered for use in the representation, transfer, access, and permanent storage of databases, textual documents, and documents containing graphics.

1.2 Impact of Framework on NARA Mission

The framework describes the scope of activities to be performed within NARA to ensure the preservation and retrieval of electronic records in a usable form.

Electronic media, such as magnetic tape and Compact Disk Read Only Memory (CD-ROM), is not permanent and has a limited shelf life. For example, under controlled environmental conditions, magnetic tape has an average shelf life of seven (7) years before the loss of data. Without maintenance, records stored on magnetic tape will be lost. New technology, such as optical disk technology, may demonstrate a greater shelf life.

Due to the continuing evolution of hardware and software, no current electronic storage media or format is expected to be sufficient to ensure permanent record retrieval. Computer hardware and software are evolving, and becoming obsolete, at a rate that may preclude record retrieval after a period of ten years, or less.

The compact nature of electronic record storage makes the task of the accessing agency and historian significantly more difficult. Records that have been adequately preserved and physically maintained may still be lost due to lack of indexing information. The density of undifferentiated records stored together in a electronic medium hinders the researcher. To be able to locate electronically stored records and record types, creating agencies and researchers must have record indexing, cross-referencing, and descriptive information available. Cross-referencing information is that which provides the direct linkage between the descriptive information and the actual archived information being described.

If this information is buried in the electronically stored record or medium itself, it cannot be considered readily available.

To avoid the loss or permanent misplacement of records stored in electronic media, NARA must define, adopt, and actively support a policy for the representation, transfer, access, and preservation of electronic records.

1.3 Task Statement

The framework presented in this document identifies a logical architecture for the representation, transfer, access, and permanent storage of electronic records to be accessioned by NARA. The requisite activities of the Federal Government agencies that create the documents to be accessioned, hereafter called the "creating agencies," is examined. The creating agencies' responsibilities in archival document preparation, scheduling, and transfer are identified as well as the responsibilities of NARA for document scheduling, accessioning, retrieval, use, and future distribution.

2 Introduction to Software Standards and Their Development

A NARA policy for the representation, transfer, access, and preservation of electronic records should be defined in terms of computing software standards. The use of software standards, defined by accredited standards organizations, provides a basis for high quality, interoperable computer systems and successful information interchange.

Software standards result from the efforts of standards organizations accredited at the Federal, national, and international levels. The term "standard" is commonly misused to refer to any computer hardware or software product that is dominant in the market at a particular time. This misconception has resulted from the predominance of these products in the marketplace. These commonly used software products are not standards; they are simply dominant products at that point in time.

2.1 De Facto and Consensus "Standards"

The predominance of particular products in the marketplace is based on the general use and popularity of these products in the commercial market at a given point in time. The "de facto" dominant products are controlled by the dominant vendor in the field, or by alliances of major vendors that aim to "corner the market" with particular software features.

In a marketplace that is governed by the fear of noninteroperability, those products that promise only partial interoperability among just one or two vendor's products can gain popularity. Even if the product only partially fulfills the vendor's promises, that product can gain a "de facto" dominance.

Products that dominate the marketplace in this way do not always offer the best quality. Such products often partially fulfill a need in the marketplace, but leave a number of other needs unfulfilled, since the planning for such products is often not comprehensive. As the computer hardware and software marketplace becomes more competitive in a particular area, however, the "de facto" dominant product can quickly lose its popularity. The apparent stability of a "de facto" dominant product is easily lost in the heat of competition.

Authorized software standards organizations, however, are not immediately tied to the rewards and fluctuations of the marketplace. In taking a longer range view, standards organizations try to plan standards more comprehensively, and try to place greater emphasis on meeting a broad spectrum of users' needs.

As standards promoting system interchange and interoperability become widely recognized, the dominance of nonconforming products tends to wane in the marketplace. Customers in the marketplace come to rely less on a particular product name, and to rely more on the conformance of software products to useful software standards. The rising popularity of the Open Systems Interconnection (OSI) standards and the OSI compliant products indicates the kind of impact that standards can have on the software industry and the marketplace.

2.2 Standards Approved by Accredited Standards Organizations

There are many accredited standards organizations operating at the Federal, national, and international levels. Accredited standards organizations that issue software standards usually rely on the work of subject area committees and subcommittees. These committees usually consist of subject area specialists representing software users, academicians, and both large and small vendors. Consensus must be reached at the subcommittee and committee level, and often approval must be given within the interested software community, before a proposed draft standard can be submitted to the parent standards organization for consideration.

2.2.1 Accredited Standards Organizations

At the Federal level, the Federal Information Processing Standards (FIPS), issued by the National Institute of Standards and Technology (NIST), require compliance in Federal government acquisition procedures within a suitable time (typically 18 months) after the standard is issued. The Department of Defense (DoD) and individual military services each issue their own software and acquisition standards for required compliance within their venues.

At the national level, the X3 Committee in association with the American National Standards Institute (ANSI) issues standards for voluntary compliance by software vendors selling products in the U.S. The Electronic Industries Association (EIA), and the associated Electronic Design Interchange Format (EDIF) organization, each approve standards useful for electronics engineering and interchange in the U.S.

At the international level, the International Organization for Standardization (ISO) and the Consultative Committee on International Telephone and Telegraph (CCITT) both offer a wide range of software standards for voluntary compliance.

2.2.2 Approved Software Standards

Most software standards are based on the highest level of currently accepted use of technology with additional structures, features, and enhancements added by standards committee specialists. Due to the requirement for consensus, most approved standards do not actually represent the forefront of the state-of-the-art. Instead, approved software standards often represent the best, well-tested knowledge available at the time.

With some exceptions, the strength of software standards is usually not in demonstrating radical advances. When a draft standard is first proposed, at that time it may represent state-of-the-art technology; by the time a standard is officially approved, however, the technology the standard is based on is often in general use. The strength of a software standard often lies in its thorough consolidation of the best features of the existing state of technology.

When a standard has been officially approved, it does not remain static. Just as "de facto" dominant products change over time due to changes in the marketplace, standards also respond to technological advances. As further requirements are recognized within the software community, standards evolve gradually through additions and modifications to the existing specifications. A few years often elapse between the initial release of a standard and the addition of such enhancements or modifications. Often these types of modifications are upward compatible with the previously specified standard.

Less often, standards may change abruptly through revolutions in state-of-the-art technology when major advancements are made.

New software standards and new software products result from these revolutionary advances.

2.3 The Importance of OSI in Current Standards

The internationally recognized Open Systems Interconnection (OSI) set of standards define a common set of data communications protocols which enable systems developed by different vendors to interoperate and which enable users of different applications on these systems to exchange information.

2.3.1 OSI Standardization

These OSI protocols were developed and adopted as standards by two international standards organizations, the International Organization for Standardization (ISO) and the Consultative Committee on International Telephone and Telegraph (CCITT).

A set of OSI standards has also been adopted by the U.S. Federal government as the Government Open Systems Interconnection Profile (GOSIP), based on agreements reached by vendors and users of computer networks participating in the National Institute of Standards and Technology (NIST) Workshop for Implementors of OSI.

2.3.2 Importance of OSI for Interoperability

The objective of OSI is to support the interconnection and interoperability of computers and systems from different manufacturers in an open systems environment.

An open system is a system capable of communicating with other open systems by virtue of implementing common international standard protocols. An open system may not, however, be accessible by all other open systems. This isolation may be achieved by physical separation or by specially designed technical capabilities based upon computer and communications security.

Both the Federal government and the private sector recognize the need to develop a set of communication protocols based on ISO's seven-layer OSI Reference Model. In the past, vendor-specific implementations of data communication protocols led to isolated domains of information, which were difficult and expensive to bridge. Recent advances in communication technology based on the OSI Reference Model offer alternatives to vendor-specific network solutions. Advances in open systems also permit the interoperations of "end systems" from different manufacturers, when required.

An "end system" contains the application processes which are the final sources and destinations of user-oriented message flows. The functions of an "end system" can be distributed among more than one computer processor.

An "end system" is differentiated from an "intermediate system," which interconnects two or more subnetworks. For example, an "intermediate system" could connect a local area network (LAN) with a wide area or packet-switching network, by performing the routing and relaying of traffic between each system. A computer processor can implement the functions of both an "end system" and an "intermediate system."

A system may provide service directly to users (acting as an "end system") and may connect subnetworks (acting as an "intermediate system") when all seven layers of OSI protocols are implemented (see paragraph 2.3.3 for list of layers). When a system performs only the functions of an "intermediate system," only the lower three layers of protocol are exercised.

A major benefit expected to result from the implementation of the OSI Reference Model is major cost reduction in the acquisition and maintenance of computer network systems. By increasing the alternative sources of supply, the standard allows users to choose competitively priced products as users are freed from dependency on products from a single vendor.

2.3.3 The Role of Protocols in OSI

In the OSI Reference Model, the communication functions are partitioned into seven layers. The seven layers, from the lowest level up, are: Physical Layer, Data Link Layer, Network Layer, Transport Layer, Session Layer, Presentation Layer, and Application Layer. Specific protocol types are designated for each layer. These seven layers, and specifications for their designated protocols and supporting functionality, are the primary structures of OSI.

The protocol layers are structured so that each layer supports the functioning of the layer above it. Thus protocol layer N provides a service to the layer above, N+1, by carrying on a conversation with layer N on another processor. The rules and conventions of the N-layer conversation are called a protocol.

An important OSI Application Layer protocol language is called Abstract Syntax Notation One (ASN.1). ASN.1 is a language primarily used for specifying data types and data values in Application Layer protocols. The Application Layer, the highest layer in the OSI model, supports the content of the information the user sends and receives. In general, the use of ASN.1 is not restricted to Application Layer protocols and may be used to specify an abstract syntax for a protocol in any layer. ASN.1 is particularly suited and widely used, however, to support the complex data types encountered in Application Layer protocols.

ASN.1 is used for defining documents in the Office Document Architecture (ODA) and Interchange Format (ODIF) family of standards, and for data interchange in the Information Resource Dictionary System (IRDS) and Remote Database Access (RDA) standards. It is expected to be incorporated into the Structured Query Language (SQL) standard within a few years. The use of ASN.1 in these standards permits users of these standards to participate in the Open System Interconnection (OSI) environment.

Because it is sufficiently robust to be useful for defining arbitrary and complex data types and values, ASN.1 is recommended by the American National Standards Institute (ANSI) Technical Subcommittee X3T2 (Data Interchange) as the basis for future ANSI standards requiring data interchange protocols.

ASN.1 was used in the development of the Document Transfer prototype software developed as part of this project. For more information on this application please refer to the documentation of the prototype system.

3 Conceptual Framework for the Exchange of Electronic Records

This section describes the structure of a conceptual framework for NARA's exchange of electronic records and descriptive information. The conceptual framework is presented as a simplified structure through which to discuss NARA policy considerations, and as a vehicle through which to represent NARA's view of information interchange. The framework and policy are discussed in a series of modules, with each module discussed separately at first and then combined to form the complete conceptual framework.

This conceptual framework is intended to provide a general, logical structure from which to view NARA's future information management and interchange activities. This generalized framework is intended to assist in identifying areas for NARA's policy and standards planning; it **is not** intended as a detailed representation of current NARA activities. A conceptual framework or architecture is very different from a physical architecture; this framework **does not** attempt to describe a physical implementation.

3.1 Record Transfer

The activities of electronic record scheduling, transfer, accession, retrieval, and dissemination are discussed in this section. The relationships among the creating agencies, NARA, and NARA record storage and retrieval are discussed conceptually.

3.1.1 Record Scheduling, Transfer, Accession, and Retrieval

This conceptual framework module describes the manner in which records are electronically transferred to NARA, reviewed, evaluated, and stored, and then retrieved and accessed by NARA personnel for internal NARA use. The structure of this framework module is illustrated in Figure 1.

Two primary modes of electronic transfer have been identified in this module: (1) the process of record scheduling, transfer, and accession from the creating agencies; and (2) the process of record handling, retrieval, and access within NARA. These two modes of electronic transfer are shown in Figure 1.

The process of scheduling records, the first step in the process, can be conducted electronically from the creating agency's location in many cases. Electronic support for the scheduling process is currently being investigated by NARA.

NARA should establish policy to support the electronic transfer of records and scheduling information from the creating agencies. Support for the actual electronic transfer of records and scheduling information from the creating agencies should be the joint responsibility of the creating agencies and NARA.

Once the scheduled records are identified by the creating agency, the agency can then electronically transfer these records to NARA. When the scheduled records are received at NARA and accessioned, the records are methodically reviewed, evaluated, and then can be electronically stored. The records must be easily accessible during the review and evaluation process, and later must be easily retrievable and accessible from storage.

The processes of both record handling and access, and record storage and retrieval depend on the availability of descriptive information about records or record groups. This descriptive information describing record content and origin, and record storage formats and location is critical to the functioning of the archival process. To support the increasing volume of electronically stored records, and the increasing volume of records stored in other media, NARA policy should establish the need to improve the methods of managing descriptive information.

Record Scheduling, Transfer, Accession, and Retrieval

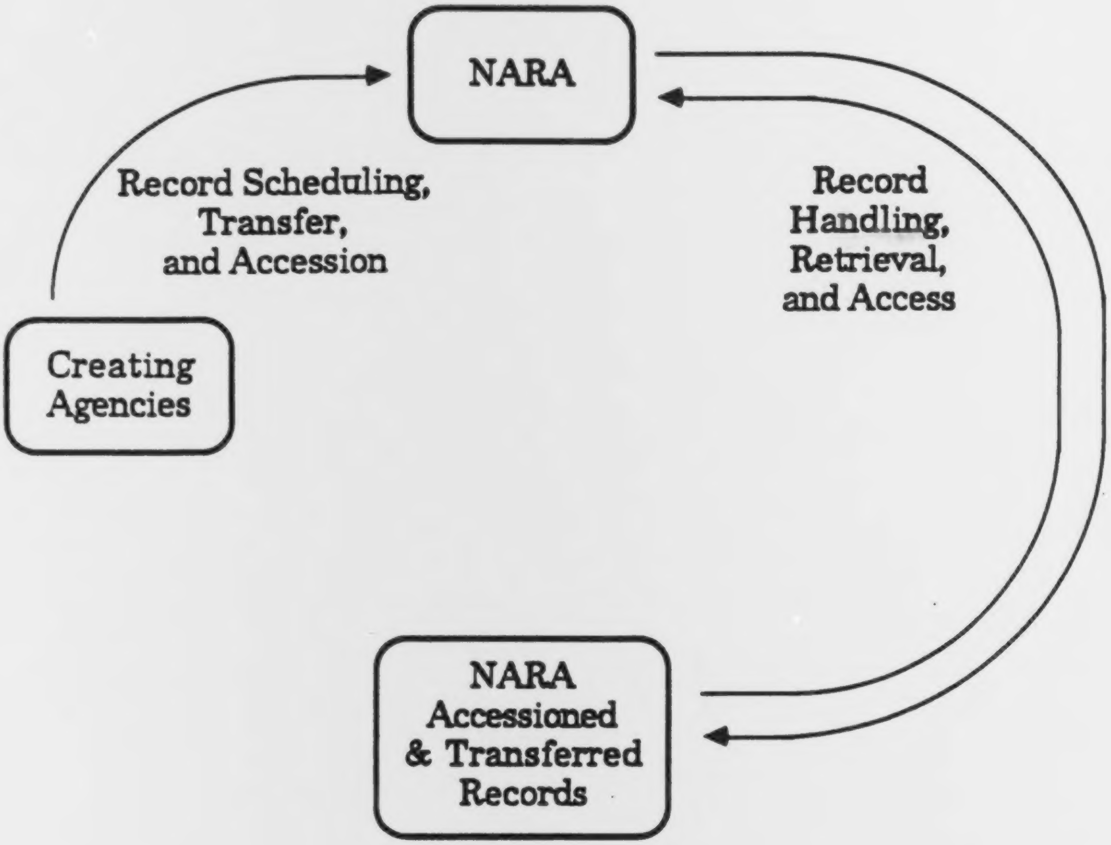

NARA
Record Scheduling, Transfer, and Accession
Record Handling, Retrieval, and Access
Creating Agencies
NARA Accessioned & Transferred Records

Figure 1

Descriptive information about records is generated: (1) during the scheduling and transfer process (i.e., often the creating agency's description of record content and origin); (2) during the review and evaluation process (i.e., NARA's detailed description of record content and origin) and; (3) during the storage process (i.e., NARA's description of record storage format, medium, and location).

Adequate data management and rapid access to this descriptive information is necessary to ensure that electronically stored records can be found and retrieved, and to ensure that meaningful descriptive information is available about record content. The capability both to retrieve records effectively within NARA, and to disseminate records to non-NARA users, depends on multiple users' easy, reliable access to complete record descriptive information. Support for descriptive information management is discussed in greater detail below.

3.1.2 Dissemination of Records to Non-NARA Users

Once the scheduled records have been transferred, reviewed, evaluated, and stored, these records must be retrievable to be available for dissemination. The previous section has discussed the process of record retrieval within NARA. Records must also be readily retrievable and available for dissemination to the creating agencies and to independent researchers.

For effective dissemination of records to Non-NARA users, descriptive record information must be readily available to non-NARA users in forming their requests. This descriptive record information would include record class identification, origin, content, cross-referencing, etc., and should refer to record storage format, storage medium, and storage location. Non-NARA users should have direct access to descriptive information on record class identification, content, and cross-referencing.

While **record information** and **descriptive record information** should be available to non-NARA users, so that users can select the desired records, **descriptive record storage information** should be invisible and inaccessible to non-NARA users. NARA policy should establish the availability of record information and descriptive record information to creating agencies and researchers. Record dissemination is illustrated in Figure 2.

In order to protect records stored electronically from intrusion, the descriptive information on record storage should be accessed by the record retrieval system, but it should not be displayed for non-NARA users. NARA policy should also address the need to maintain descriptive record storage information without public access.

Dissemination of Records to Non-NARA Users

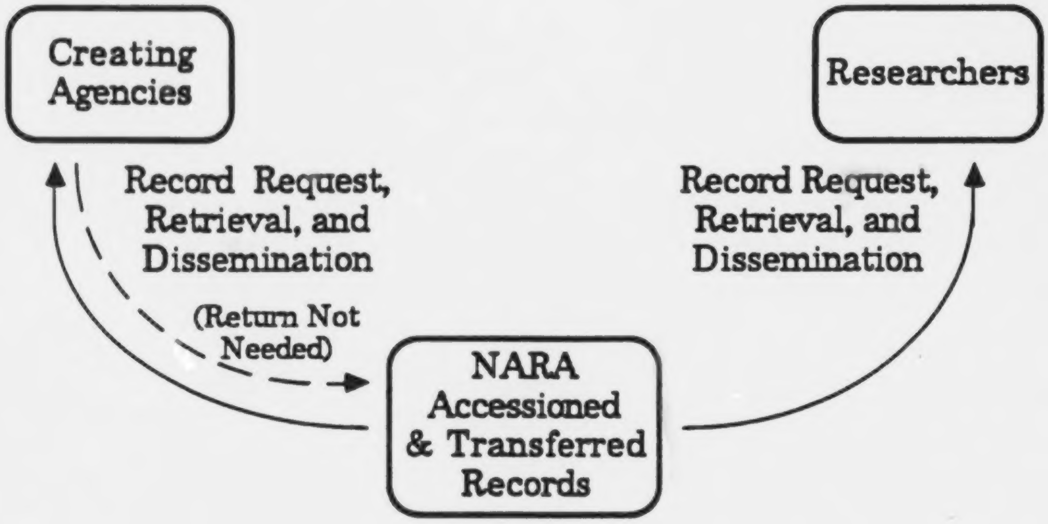

Creating Agencies
Researchers
Record Request, Retrieval, and Dissemination
Record Request, Retrieval, and Dissemination
(Return Not Needed)
NARA Accessioned & Transferred Records

Figure 2

NARA policy should establish plans for the computerized management and dissemination of records. The dissemination of records should be accomplished through a computer system that can validate user identity and user class, can support the user in selecting records, can support NARA in locating and retrieving stored records, and can support NARA by sending records electronically to users.

NARA policy should establish the need for implementing an automated system for record interchange, supporting record transfer both to and from NARA, and should define the need for requiring non-NARA user registration into this automated system. User registration will provide the record interchange system with information to identify valid non-NARA users. This automated interchange system should include both a computer system, with appropriate data management software, and electronic media for data interchange, such as access to an electronic network. NARA policy should define the functions of such an automated record interchange system. These functions should include features such as the user registration function, the record accession function, the record request function, and the record transfer function. When user registration is implemented on an automated system supporting record interchange, the system can be used both in the accessioning and transfer of records from creating agencies, and to support system users in information requests and record transfer.

NARA policy should address system user registration issues. First, valid non-NARA users should provide registration and address information to NARA, for NARA personnel to verify and input as descriptive information to the supporting system. This registration should occur **once per user**, or once per user organization, and should be modified only if the user changes affiliation or network address, etc. Once the user information has been input to the computer system, which has been appropriately programmed, the system should thereafter be able to validate electronically user requests for records.

When a non-NARA user accesses the record interchange system to request records (by selecting descriptive record information), then the computer system supporting the descriptive information should electronically validate the user's identity as to the type, affiliation, and location of the user. Once the user's identity has been validated, then the user can complete the query. For the non-NARA user the record storage information should be retrieved by the system, but not displayed to the user.

At this point, the computer system can then either retrieve the requested records directly, if the records are stored on-line, or can issue a request for the selected records, indicating to NARA personnel the records' storage information. As the record request is being processed, the computer system should issue a message to the user that the record request has been received and validated, and that the record will be issued within a designated span of time.

Once the requested records have been retrieved, either electronically or manually, these records should be electronically transferred to the requesting agency or researcher.

Since record retrieval and transfer involves only electronic **copies** of the electronically stored records, the integrity of the originally stored records cannot be compromised by the record transfer. Since no original records will be sent, records **will not** be returned from the requesting/creating agencies, so **no changes** can be introduced by the agencies during an examination of records transferred in this manner.

With this policy and the implementation of these automated record interchange procedures, NARA will be able to determine unequivocally, and enforce, that the records stored at NARA will remain as they were originally received from the creating agencies.

3.1.3 Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination

The complete Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination is illustrated in Figure 3. This framework section combines the previously discussed framework modules. The procedures demonstrated in this combined framework module are: (1) record scheduling, transfer, and accessioning from the creating agencies and record transfer to NARA; (2) record handling, storage, and retrieval within NARA; and (3) record requests, searches, retrieval, and dissemination to non-NARA users such as the creating agencies and researchers.

NARA policy should establish the requirements, as described in the previous sections, to provide automated support to NARA and the creating agencies in record scheduling and transfer. NARA policy should also establish requirements, as described above, to provide automated support to creating agencies, researchers, and NARA personnel in the selection, retrieval, and transfer of the appropriate accessioned records.

This automated support should include a computer system, with appropriate data management software, and access to an electronic data facility, such as an electronic network. For such an automated system to operate, NARA policy must place emphasis on the management and support of descriptive record information, which is critical to successful operation of the

Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination

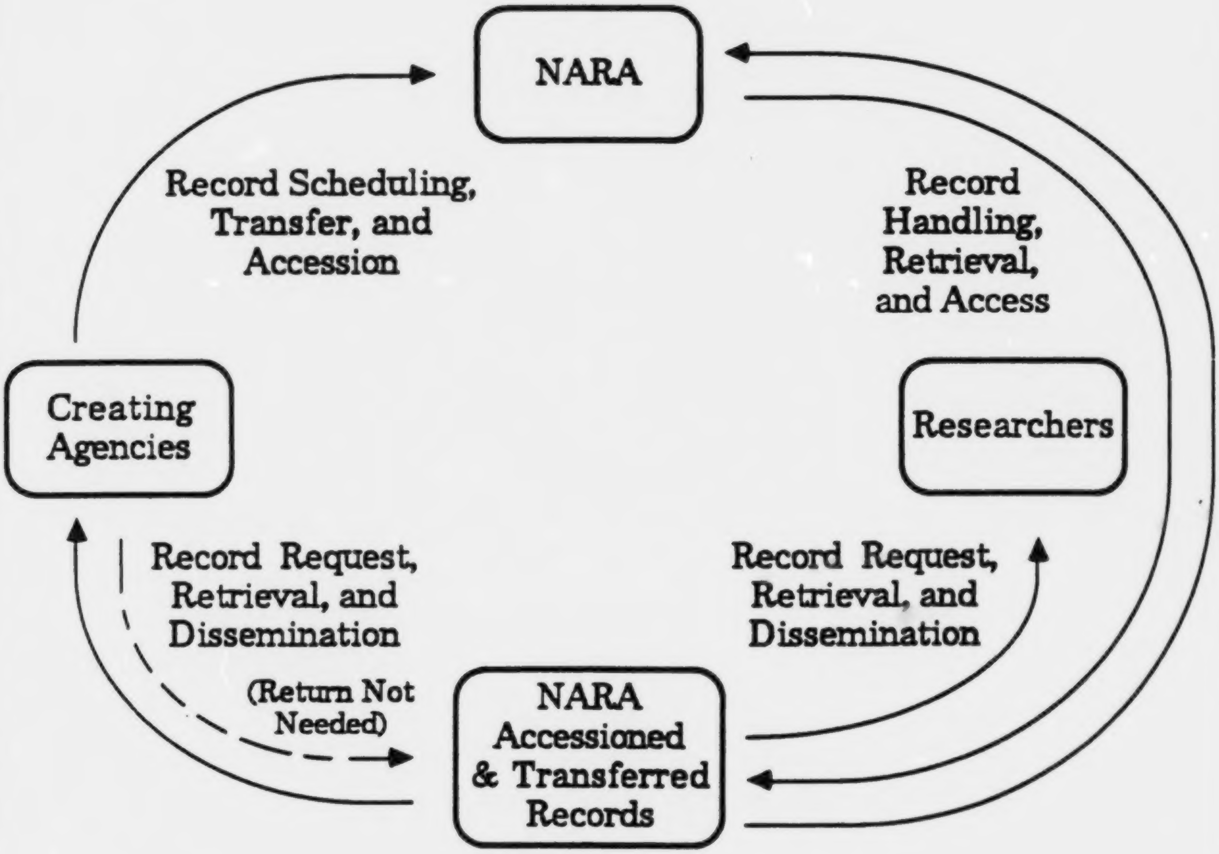

NARA
Record Scheduling, Transfer, and Accession
Record Handling, Retrieval, and Access
Creating Agencies
Researchers
Record Request, Retrieval, and Dissemination
(Return Not Needed)
Record Request, Retrieval, and Dissemination
NARA Accessioned & Transferred Records

Figure 3

electronic interchange system described.

NARA policy should designate automated support for all descriptive record information and for all descriptive record storage information. NARA policy should require the automation of this descriptive information in a manner integrated with the electronic interchange system. This integrated automation of both subject information (i.e., records) and descriptive information (i.e., about records) will permit the electronic interchange system to operate, and will simplify the handling of records, both within and outside NARA.

3.2 Descriptive Information Capture and Interchange

This section discusses the creation, capture, management, and interchange of record scheduling and record descriptive information generated during the record handling process. As stated previously, descriptive information about records is generated: (1) during the scheduling and transfer process (i.e., often the creating agency's description of record content and origin); (2) during the review and evaluation process (i.e., NARA's detailed description of record content and origin); and, (3) during the storage process (i.e., NARA's description of record storage format, medium, and location).

3.2.1 Transfer of Scheduling and Descriptive Information from Creating Agencies

Scheduling and other descriptive information about records is produced by the creating agencies, in conjunction with NARA, to be transferred to NARA along with the specified records. Once received by NARA, this scheduling and descriptive information is reviewed and evaluated by NARA staff, who may augment this information as necessary. After this scheduling and descriptive information is reviewed, it is stored to maintain information about record content and structure, and to maintain information for record access and retrieval. These activities are depicted in Figure 4.

3.2.1.1 An Example Problem: Database Transfer

The importance of descriptive information about record content and structure is particularly obvious in the process of exchanging database records. One of the difficulties encountered by NARA is the lack of an adequate means of receiving and transferring database information without loss of meaning. At this time, meaningful database information can only be adequately transferred: (1) in the limited format of reports and query results, through query languages such as SQL; (2) from active database to active

Transfer of Scheduling and Descriptive Information to NARA from Creating Agencies

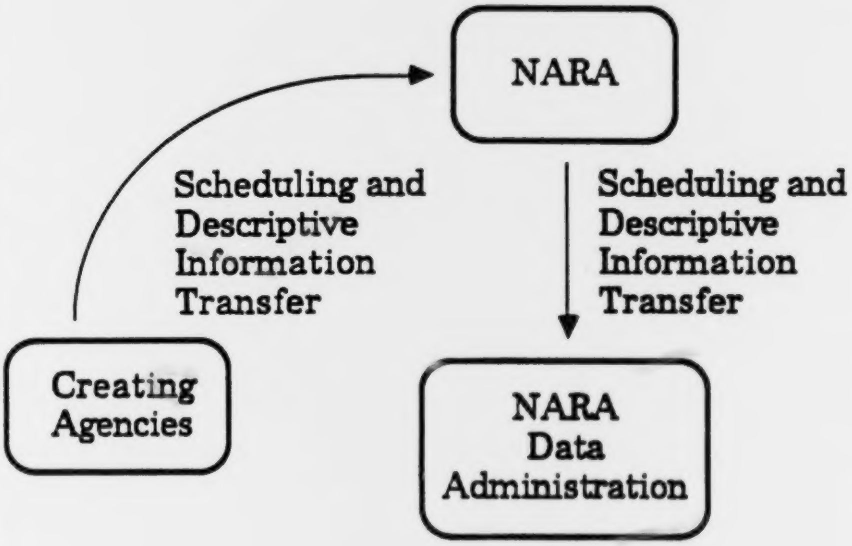

NARA
Scheduling and Descriptive Information Transfer
Scheduling and Descriptive Information Transfer
Creating Agencies
NARA Data Administration

Figure 4

database (i.e., where active database denotes the presence of both computer hardware and database management system (DBMS) software); or, (3) when database information is represented in a flat file of data, and is accompanied by complete descriptive information of the structure and other features of the original database.

The first option provides accurate but incomplete information about a database, while the second option is prohibitively expensive and unwieldy with such an array of database software and hardware available. At the present time, the third option is the best choice for the representation and exchange of complete database record information.

If an agency is sending NARA a "dump" of the contents of a database in a flat file format, for example, a record containing only this data flat file does not indicate the structure of the database. Since the **use** of the data in the database is dependent on the **structure** of the database, as well as the structure of the reports and queries to which the database was designed to respond, such a flat file record is meaningless. While the data may, technically, be considered to be captured, the meaning and significance of the data and the database is lost.

3.2.1.2 An Example Solution: Database Descriptive Information

In order to restore meaning to a flat file "data dump" of a database, descriptive information about the schema structure of the database, as well as about its report and query structures, must be captured, transferred along with the data flat file, and maintained in conjunction with the data flat file. Capture, transfer, and coordinated maintenance of both the flat file and complete descriptive information is imperative to subsequent understanding of the record database and its data.

The data flat file and the descriptive information defining the database structure must be maintained and retrieved together to provide meaningful information on the source database. Since these parts of the same record will usually be electronically stored separately, because they are different types of data and have to be accessed differently, NARA's maintenance of descriptive cross-referencing information (e.g., in this case linking the flat file part of the record to the descriptive part of the record) will be critical to the later retrieval and use of electronically stored records.

3.2.1.3 NARA Policy Needed for a Coordinated NARA-wide Data Administration Function

Easy access to this scheduling and descriptive information is critical to the functioning of the record retrieval and interchange system, and to the functioning of NARA's record review and evaluation procedures. To ensure easy access and coordination of this scheduling and descriptive information, this information should be managed by a centralized NARA Data Administration function integrated across all NARA offices. The various Data Administration functional areas can retain separate areas of responsibility, but should be closely coordinated through a centralized Data Administration function.

NARA policy should establish the need for an integrated NARA-wide Data Administration function, and should describe procedures appropriate to implement this function in the near future.

3.2.1.4 NARA Policy Needed for Accessible Descriptive Information within NARA

Such a Data Administration function should be responsible not only for **monitoring the formats of descriptive information**, but should also be responsible for ensuring the **easy availability of the descriptive information content** to all interested users, within and outside NARA. In order to achieve this availability goal, the Data Administration function must emphasize the use of standards. NARA policy should ensure that such an integrated Data Administration function will be supported by an integrated, automated system so that the Data Administration function can fulfill this mission.

NARA policy should establish the need to electronically capture this scheduling and descriptive information at the points at which it is created, at the creating agencies and within NARA. NARA policy should establish the need to store this descriptive information on an automated system, as soon as the information is created, so that it can be simultaneously available to multiple system users within and outside of NARA.

The need for an automated system to provide adequate support to such an integrated Data Administration function is discussed below in the section, "NARA Policy for Data Administration Support."

3.2.2 Internal Interchange of Scheduling and Descriptive Information within NARA

Records and descriptive record information should be readily available to NARA personnel users during: (1) the process of record scheduling, accessioning, handling, review, and evaluation; and (2) the stage when records are stored, either electronically or in other media. As described previously, descriptive information about records plays a critical role in the reliable and ready access to these records.

An integrated NARA-wide Data Administration function, as previously described, is central to the successful, rapid interchange of scheduling and descriptive information within NARA. The support of an integrated, automated system, as previously described, is critical to the success of such a centralized Data Administration function, and is imperative to the full interchange of scheduling and descriptive information within NARA (as well as for information interchange with other agencies).

3.2.2.1 Internal Information Availability

The purpose of such complete internal availability of scheduling and descriptive information within NARA is: (1) to speed up the procedures of NARA record handling, review, and evaluation; and (2) to enable NARA personnel, and possibly designated creating agency personnel, to recommend appropriate record cross-referencing information among a variety of record categories and classes.
NARA policy should direct the integrated, NARA-wide Data Administration function to acquire and implement integrated automated support for Data Administration of descriptive information, that will permit multiple NARA system users to simultaneously access records and descriptive information. The type of software which should be utilized is typically referred to by several different names such as Data Dictionary, Information Resource Dictionary, Data Directory, etc. The term used in this document shall be Data Dictionary. A Data Dictionary is a software package that is designed for the specific purpose of documenting and retrieving information about the various characteristics of an organization's data assets. These documented characteristics could include such things as the data asset's name, a description, the size, the location, the originator, the subject area, etc. For a more complete discussion of Data Dictionary the reader should refer to any of a number of popular publications on the subject, such as The Data Dictionary Concepts and Uses by Charles J. Wertz (published by QED Information Sciences, Inc.).

The simultaneous use throughout NARA of such integrated, automated descriptive record information is indicated, in summary, in Figure 5. The top box indicates the multiple areas of NARA that receive records, as well as scheduling and descriptive information, from the creating agencies. In handling and reviewing records, these areas of NARA should not only enter all descriptive record information directly into the Data Administration system, NARA personnel should also use this system to query information on the status of both currently and previously handled records.

Internal Interchange of Scheduling and Descriptive Information within NARA

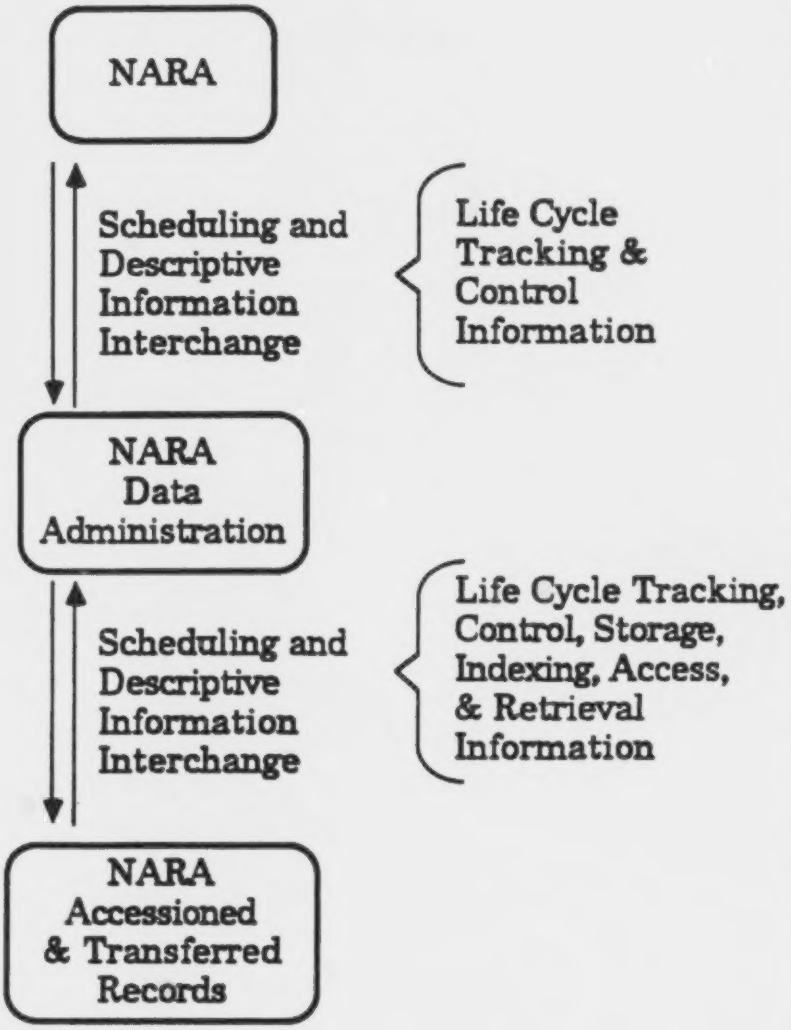

NARA
Scheduling and Descriptive Information Interchange
Life Cycle Tracking & Control Information
NARA Data Administration
Scheduling and Descriptive Information Interchange
Life Cycle Tracking, Control, Storage, Indexing, Access, & Retrieval Information
NARA Accessioned & Transferred Records

Figure 5

The middle box indicates the NARA-wide coordinated Data Administration function that monitors and provides automated support for NARA's descriptive record information. Through the Data Administration function, NARA personnel should be able to access the following information when records are being accessioned: (1) the content and status of incoming descriptive record information, as it is entered into the data dictionary system, and (2) the descriptive information in the data dictionary system about other currently handled and previously accessioned records.

The coordinated Data Administration function should continue and expand the current information life cycle tracking and control effort, by providing automated support for this information via the data dictionary system. This automated information life cycle tracking and control information should be fully integrated in the same data dictionary system with all the other descriptive record information, such as record scheduling, transfer, electronic storage, indexing, cross-referencing, access, and retrieval descriptive information.

The last box indicates the storage location of NARA accessioned and transferred records, which should be able both to receive and transfer descriptive information within NARA.

All descriptive record and life cycle tracking information should be available to authorized NARA personnel on a continuous basis from the data dictionary system under the direction of the coordinated Data Administration function.

3.2.2.2 Data Dictionary Administration Support

A Data Dictionary Administration function should be organized under the direction of the centralized Data Administration function. The director of the coordinated Data Administration function should expect to hire personnel trained to perform the technical data dictionary administration procedures and trained to physically operate the data dictionary system. These Data Dictionary Administration personnel should structure and operate the NARA-wide integrated data dictionary system in accordance with NARA centralized Data Administration goals and in accordance with stated NARA policy.

3.2.2.3 Limits to Internal Information Creation and Modification

It should be noted that simultaneous access to records and descriptive information does **not mean** that multiple system users will be able to modify this information simultaneously.

A data dictionary system, as part of the NARA automated system, should permit the centralized Data Administration function to designate which particular NARA personnel have responsibility for, and are therefore authorized to modify or add to, particular types of descriptive information about records.

Through the use of security features of the data dictionary system, all records can be designated as unmodifiable, and only specially designated users should be permitted the capability to modify particular types of descriptive record material.

The status of such authorized individuals should be defined and maintained in the automated system by the centralized Data Administration function. The automated system should be able to validate a user's authorized permission status through the identifying characteristics of the person's system "login" and individual password.

Only one authorized user should be able to modify a file at any point in time. Multiple users throughout NARA, however, should have the capability to access, review, and note the status of records and descriptive information in the process of evaluation, or records that have been stored in various media.

3.2.2.4 NARA Policy Needed for Data Administration Support

NARA policy should establish that, the integrated, automated support for the Data Administration function should include both database management system (DBMS) software and data dictionary system (i.e., repository) software in order to successfully manage both records and descriptive information.

A NARA software acquisition policy should be defined requiring that both DBMS and data dictionary system software used by NARA should comply with appropriate standards. The data dictionary system software should comply with the Information Resource Dictionary System (IRDS) standard. If a relational DBMS is used, its query language should comply with the Structured Query Language (SQL) standard.

3.2.3 Queries for Record Access Via Data Administration System

The record scheduling and descriptive information provided by the creating agencies to NARA has been shown to provide NARA with the basic information to be able to then review, describe, store, cross-reference, and retrieve these records.

Some of this descriptive information is appropriate only for use within NARA itself, such as the exact location and format of electronic storage of particular types of records. Other types of descriptive information are useful and appropriate for access by the creating agencies and authorized researchers.

Authorized creating agencies and researchers, identified by designated user passwords or codes, should be granted access to some descriptive record information, such as scheduling, cross-referencing, and retrieval information. **Read-only** privileges should be granted to these non-NARA users, so that no records or descriptive information can be modified. These authorized non-NARA users should, however, be provided the capability to search for specified types of records (as permitted in their areas of authorization), and should be able to retrieve such records in an automated manner. Figure 6 illustrates this activity.

This support for record access from non-NARA users should be provided by the centralized NARA Data Administration function, and its supporting automated data dictionary system. If automated correctly, this activity should place only a minor burden on the Data Administration function.

The automated data dictionary system, which the Data Administration function should already be tasked to maintain, will provide most of the support needed for these non-NARA queries. If the descriptive information about records is maintained by NARA, as discussed in previous sections, the extension of the use of this information to non-NARA users requires only the additional registration of authorized agencies/researchers into the automated system.

NARA policy should address the need to provide accessible, non-modifiable descriptive information to authorized non-NARA users. This descriptive record information will provide creating agencies and researchers with the capability to be able to select the record types that they wish to access, so that these non-NARA users can then submit requests for record retrieval. Both the descriptive record information and the particular record groups should be specified in policy as available for retrieval in an automated manner. NARA policy should also specify that both records and their descriptive information should be protected from modification by these non-NARA users.

3.2.4 Descriptive Information Interchange Framework

This framework, depicted in Figure 7, illustrates how the descriptive record information, received from the creating agencies and generated within NARA, should be widely accessible for use both by authorized NARA users and authorized non-NARA users. This descriptive record information should provide the users of the automated data dictionary system with information about the record group origins, identifying features, general content, cross-

Queries for Record Access Via NARA Data Administration System

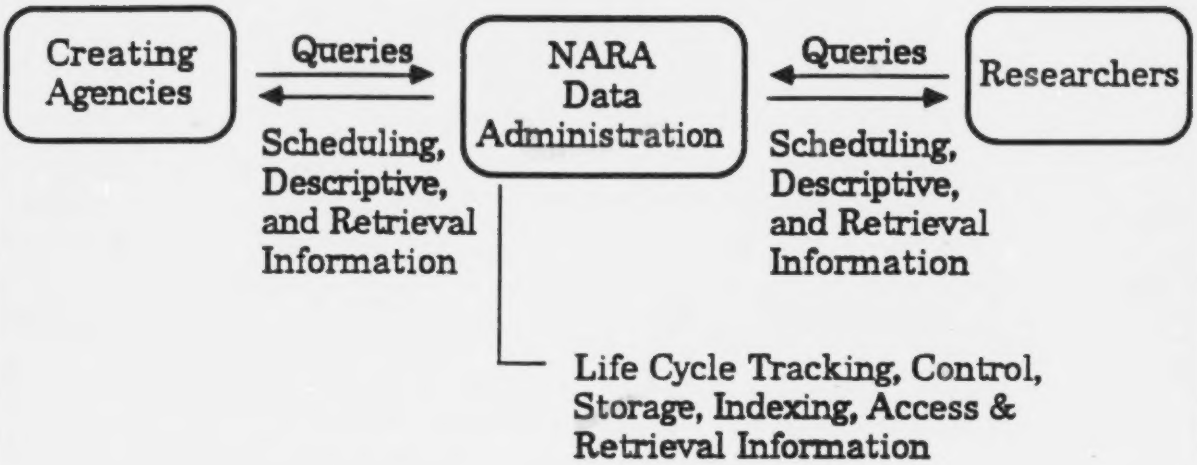

Creating Agencies
Queries
Scheduling, Descriptive, and Retrieval Information
NARA Data Administration
Queries
Researchers
Scheduling, Descriptive, and Retrieval Information
Life Cycle Tracking, Control, Storage, Indexing, Access & Retrieval Information

Figure 6

Descriptive Information Interchange Framework

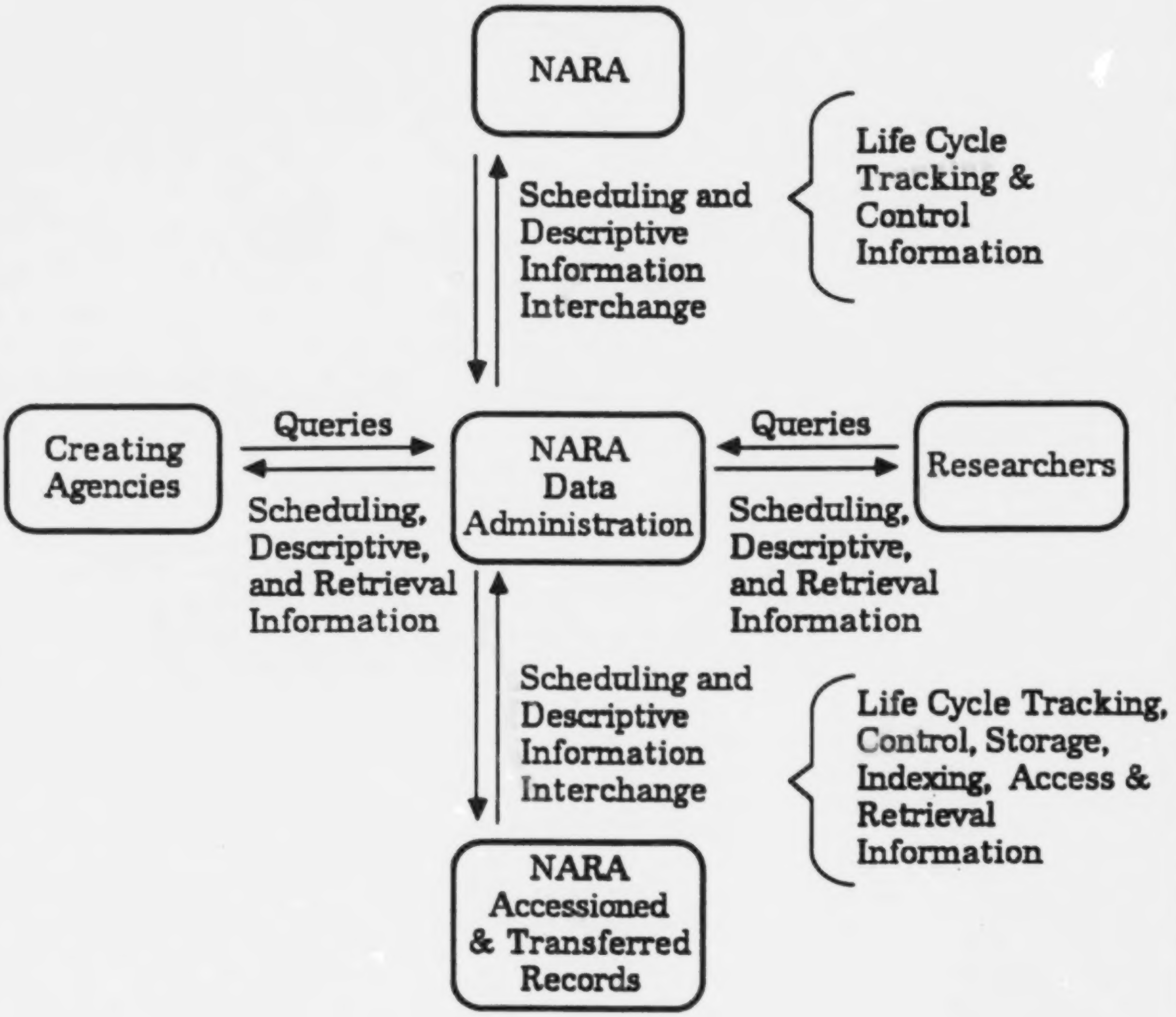

NARA
Scheduling and Descriptive Information Interchange
Life Cycle Tracking & Control Information
Creating Agencies
Queries
Scheduling, Descriptive, and Retrieval Information
NARA Data Administration
Queries
Researchers
Scheduling, Descriptive, and Retrieval Information
Scheduling and Descriptive Information Interchange
Life Cycle Tracking, Control, Storage, Indexing, Access & Retrieval Information
NARA Accessioned & Transferred Records

Figure 7

-referencing, and access/retrieval information for specifiable record groups. With this descriptive information at hand, both NARA and non-NARA users should be able to retrieve particular record groups, according to the type of the users' authorization.

3.2.4.1 NARA Policy Needed for Capture of Scheduling Information Directly from the Creating Agencies

During the record scheduling and accessioning process, the creating agency, in collaboration with NARA, creates descriptions of record content and origin according to the agency's perspective of the records. Since many creating agencies will later wish to access records from NARA, these agencies will benefit by providing the most accurate and complete description of record content and origin as feasible, so that record retrieval will be possible at a later time. NARA policy should establish the need to capture scheduling information in automated form directly from the creating agencies, whenever possible, and to store this descriptive information in an integrated, automated system.

3.2.4.2 NARA Policy Needed for Cross-Referencing Information

During NARA's record review and evaluation process, NARA personnel may further refine these descriptions of record content, and may assign additional descriptive information to the record. When either NARA personnel or creating agencies are aware of cross-referencing information appropriate to particular records, it is at this point that this cross-referencing descriptive information is assigned to these records.

Through the use of the Data Administration's data dictionary system, both NARA personnel and creating agencies' personnel should have access to descriptive information about previously accessioned records. Related descriptive records should be reviewed as new records are received to establish appropriate cross-referencing links.

NARA policy should establish the need to capture NARA's record review and evaluation information at the time of composition, and to store this descriptive information in an integrated, automated system.

3.2.4.3 NARA Policy needed for Electronic Record Storage Information

During the storage process, when NARA personnel are assigning records to be stored in particular formats, in particular electronic (or paper) media, and in particular storage locations, descriptive record storage information should be assigned to particular records. NARA policy should establish the requirement to capture NARA's record storage information, particularly electronic record storage information (such as location, media, and format), at the time of record storage, and to store this descriptive information in an integrated, automated system.

To ensure adequate control and accessibility of this descriptive information, NARA should automate this information on a standardized data dictionary system, as previously discussed.

3.3 Framework for Record Accession/Transfer and Descriptive Information Interchange

The unified NARA framework of Figure 8 depicts, in summary form, both NARA's internal and external information interchange requirements, and NARA's record and descriptive record information interchange requirements.

Three detailed summary frameworks are combined in this unified framework. The unified framework of Figure 8 combines the features of Figure 3, "Framework for NARA Record Scheduling, Transfer, Request, Retrieval, and Dissemination," Figure 4, "Transfer of Scheduling and Descriptive Information to NARA from Creating Agencies," and Figure 7, "Descriptive Information Interchange Framework."

Creating agencies send both records and descriptive record information to the appropriate group within NARA for record handling, review, accessioning, and storage. Records and descriptive record information should be electronically transferred directly from the creating agency, in most cases. Within NARA, record handling, review, accessioning, and storage should also be supported by an automated system and electronic information interchange.

Scheduling and descriptive information should be sent, in an automated manner, to the centralized NARA Data Administration function as soon as it has been received and reviewed for accuracy. This same information should be exchanged with the site at which the accessioned and transferred records are stored, so that this descriptive information can be incorporated into and integrated with record storage information.

Record handling, review, cross-referencing, and other access information should accompany the newly received records within NARA to the site of accessioning and storage, and should be entered into the NARA Data Administration data dictionary system for the immediate accessibility of this information throughout NARA.

Record Transfer/Accession and Descriptive Information Interchange Framework

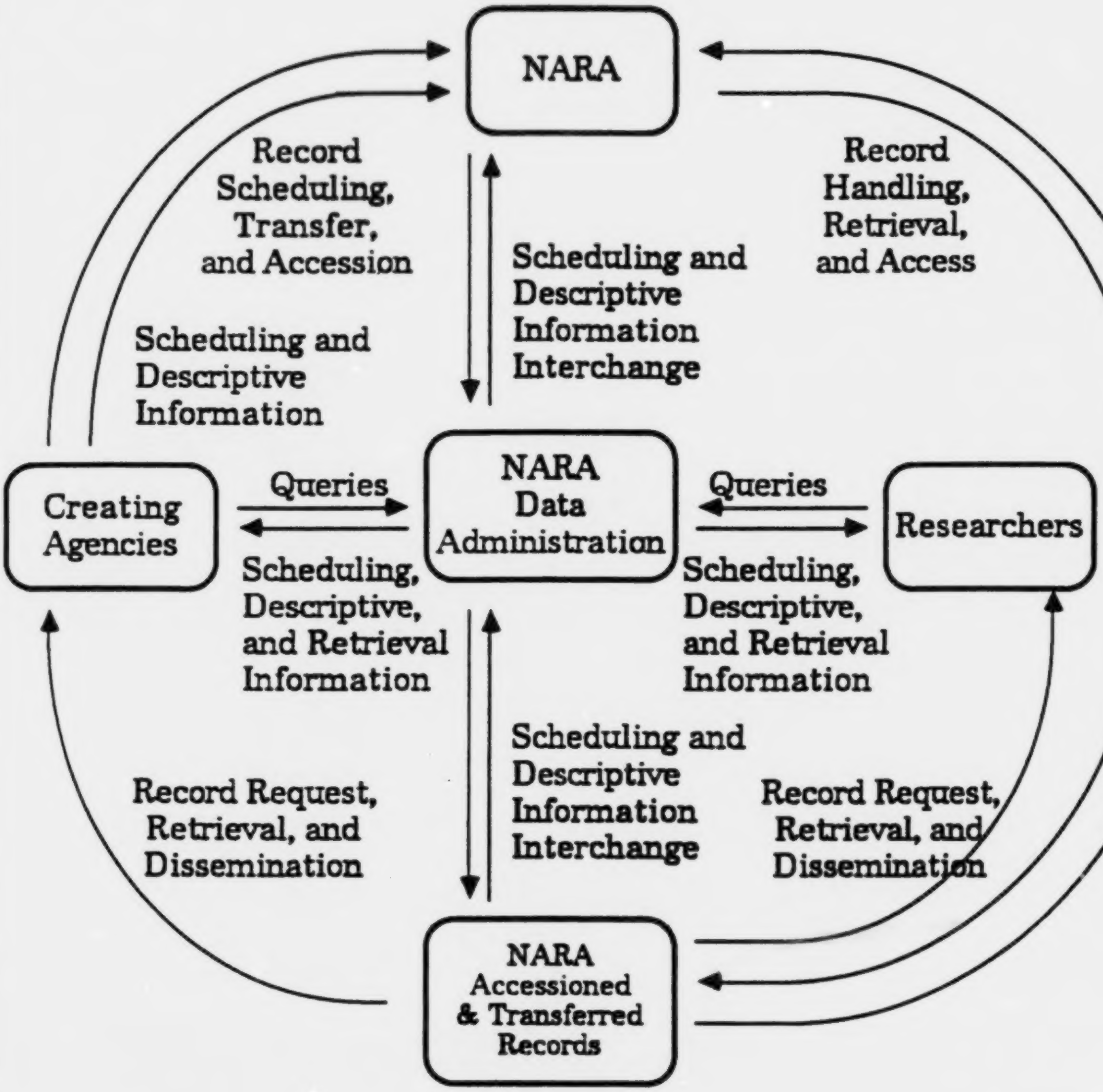

NARA
Record Scheduling, Transfer, and Accession
Scheduling and Descriptive Information
Scheduling and Descriptive Information Interchange
Record Handling, Retrieval, and Access
Creating Agencies
Queries
NARA Data Administration
Queries
Researchers
Scheduling, Descriptive, and Retrieval Information
Scheduling, Descriptive, and Retrieval Information
Scheduling and Descriptive Information Interchange
Record Request, Retrieval, and Dissemination
Record Request, Retrieval, and Dissemination
NARA Accessioned & Transferred Records

Figure 8

Record handling and descriptive information handling should be captured and monitored in the NARA Data Administration data dictionary system, to maintain information life cycle tracking and information interchange. The creation and maintenance of information life cycle tracking, as part of the integrated Data Administration activities, is shown and discussed in greater detail in conjunction with Figure 5.

Since it will receive queries for topical information from authorized creating agencies and researchers, the NARA Data Administration data dictionary system should be structured to provide descriptive record information on the types of accessioned record groups that may satisfy these topical information requests. From this descriptive record information, these non-NARA users should be able to formulate requests for the particular record groups in which they are interested.

The record request could be placed directly with the data dictionary system's integrated user interface, or it could be placed separately through the record storage site, if known to the non-NARA user. The record request could result in both the retrieval (either automated or manual), and in the automated, electronic transfer and dissemination of selected record groups from NARA to the creating agencies and researchers. Since the original electronic record will never be sent to the user, but only an electronic copy, there is no need for record return, and no possibility for record modification by non-NARA users.

4 Creating Agency's Framework View

This section discusses the perspective and responsibilities of the creating agencies in their provision of records, scheduling information, and other descriptive record information to NARA in electronic form.

4.1 NARA Policy Needed for Shared Responsibility for Electronic Interchange

NARA should establish a policy to support the electronic interchange of records and scheduling information with the creating agencies. To aid in this process, NARA policy should direct funding for the development and distribution of computer-aided support to aid outside agencies in providing scheduling, descriptive record information, and records directly to NARA in electronic form. This same computer-aided support should be capable of aiding the creating agencies in requesting previously accessioned records and record information from NARA.

Support for the electronic transfer of records and scheduling information from the creating agencies should be the joint responsibility of the creating agencies and NARA. Both NARA and the creating agencies should contribute funding for, and cooperate in, planning and developing appropriate computer-aided support for information interchange between NARA and the other agencies.

Once NARA has provided the software and demonstrated a prototype implementation of such a system, to demonstrate the system's capabilities to the agencies, the individual agencies should provide funding to support the acquisition of the hardware needed at their site to support the NARA-provided software.

This cooperative plan ensures that NARA will be able to direct the project, with guidance from the creating agencies. In addition, this plan ensures that the creating agencies expecting to benefit most from the use of such an electronic information interchange system will participate in the early stages of system implementation. Once this system is operational at NARA and in some creating agencies, other agencies are expected to be convinced by the utility of such system demonstrations. A number of these other agencies are expected to decide to acquire this same system at that time.

Working together, NARA and the agencies should share the responsibility of supporting the electronic interchange of records, scheduling information, and other descriptive record information between NARA and other Federal government agencies.

4.2 NARA Policy Needed for Record and Descriptive Information Format and Media

NARA policy for electronic interchange policy should define guidelines for the use of electronic interchange and storage formats and media as most appropriate for particular types of records.

4.2.1 Guidelines Needed for Descriptive Information Received from Creating Agencies

NARA policy should define guidelines for the acquisition of scheduling and other descriptive record information received from creating agencies. This policy should address not only the electronic interchange of this information, but also its content.

NARA receives much record scheduling and descriptive record information from the creating agencies. Unfortunately, this scheduling and descriptive information is often far from accurate. While NARA does not, at this time, have the authority to require that creating agencies provide accurate information, NARA policy should nevertheless address this problematic area.

Part of the problem with receiving accurate descriptive record information from the creating agencies may result from the lack of experience of creating agency personnel with the record material being scheduled. This problem could possibly be addressed by a NARA policy encouraging creating agencies to involve agency project area personnel in the scheduling process in order to provide guidance to the creating agency personnel tasked with providing scheduling information to NARA.

NARA policy could also establish guidelines for NARA regularly to report gross record scheduling errors back to the creating agency, to the appropriate supervisory personnel. Creating agencies are more likely to attempt to address record scheduling errors if they are able to see the full extent of the errors generated by their agency. Creating agencies cannot know the extent of their record scheduling errors unless either they discover these errors through subsequent record queries, or if they are informed by NARA. The tone of this reporting of scheduling errors to the other agencies should be informative, and not accusatory, in order for agencies to be inclined to correct these information errors.

A newsletter sent out to the creating agencies, written with a sense of humor as well as with the purpose of delivering a message, could point out the more extreme errors in a humorous way, and could also point out NARA progress with other agencies in implementing an electronic information interchange system. Such a newsletter could serve a number of purposes, including informing creating agencies of: (1) NARA policy regarding the format, content, transfer, and handling of their records; (2) gross discrepancies between the actual records and the record scheduling information provided by the creating agencies and, (3) information for the agencies about NARA progress in establishing an automated system for record, scheduling, and descriptive information interchange.

4.3 Responsibilities of the Creating Agencies

Future record access by the creating agency and researchers depends on the creating agency's compliance with NARA guidelines concerning record scheduling and descriptive information when the records are submitted to NARA.

4.3.1 Creating Agency Access to Accessioned Records

In instances where creating agencies often reference their own previously accessioned records, these agencies understand the value of providing accurate record scheduling information, as this benefits their own search, query, and record retrieval process. In instances where the creating agencies rarely reference their own previously accessioned records, however, correct record scheduling information is not valued and is, therefore, often not provided to NARA.

NARA cannot take responsibility alone for the verification of correct record scheduling information, record interchange, correct record access, and correct record retrieval. NARA does not have the resources to verify the scheduling information for each record group received from a creating agency; neither does NARA have the resources or authority to require and police creating agencies to provide adequate record scheduling information.

4.3.2 Loss of Accessioned Records through Misidentification

In order for accessioned records to be accessible and retrievable, the scheduling and descriptive record information accompanying the record groups **must be correct**. When the scheduling and descriptive record information received from the creating agency is incorrect, the record is essentially **lost**. Even though the record itself is actually electronically stored (or stored in another manner) and maintained, access and retrieval of any meaningful information in the record is effectively destroyed because the record is misidentified.

The creating agencies must take responsibility for providing complete and correct record scheduling information about records being accessioned by NARA. While NARA can catch isolated incidents of record scheduling misidentification, NARA cannot be expected to verify each record schedule against each record group. It must be up to the creating agencies to take responsibility for record schedule verification.

4.3.3 Role of Data Dictionary System in Record Schedule Checking

While electronic media can speed the process of record and descriptive information interchange, and provide easier access to records where correct record scheduling and description has been maintained, electronic media cannot verify correct record scheduling information from the creating agencies. If creating agencies provide scheduling information to NARA through an automated data dictionary system with an integrated system user interface, the data dictionary system can be used to screen and check for likely **types** of scheduling information responses that agencies may provide. Only the creating agency itself, however, can fully verify the accuracy of record scheduling information.

4.3.4 Creating Agency Responsibilities for Accurate Information Interchange

The software for the information interchange system, to be provided by NARA to the creating agencies, should support the capture of adequate information on: (1) record scheduling and description; (2) electronic record storage format; and, (3) electronic storage media. It is the responsibility of the creating agencies to provide guidance to NARA on the design of such a system, and to utilize the developed system to provide accurate descriptive record information to NARA.

4.3.4.1 Creating Agency Participation in NARA's Information Interchange System

Creating agencies should seek information about and participate in (or plan for future participation in) NARA's interagency information interchange system. If a creating agency cannot participate in installing the system on-site, the agency should continue to remain informed on the status of the information interchange system, and should continue to evaluate future possible participation in the system.

4.3.4.2 Provision of Adequate Record Schedule and Descriptive Record Information

Creating agencies should bear the responsibility for verifying the accuracy of their own record scheduling and other descriptive information before it is released to NARA. The record scheduling information should follow the guidelines provided by NARA. If an agency disagrees with NARA's record scheduling guidelines, it is the responsibility of the creating agency to report this dissatisfaction to NARA and to negotiate an acceptable arrangement with NARA for providing such information.

4.3.4.3 Provision of Records in Acceptable Electronic Record Formats

Creating agencies should take the responsibility for following NARA policy by providing electronically stored records in formats that NARA can handle. A format can be considered to be a document representation format, a database representation format, a data interchange format, etc.

If an agency sends an electronically stored record to NARA in a format that deviates from NARA's policy, NARA may not have the resources or capability to copy this unusual format into one that NARA can maintain over time and retrieve later. If NARA cannot maintain or retrieve the record or record description provided, the record is effectively lost and unaccessible in the future as well as in the present. When an agency sends records to NARA in unacceptable formats, the agency is, in effect, throwing the record(s) away.

4.3.4.4 Provision of Records in Acceptable Electronic Media

Creating agencies should take the responsibility for following NARA policy by providing electronically stored records in acceptable media that NARA can handle. Electronic media can be considered to be magnetic tape, optical disks, floppy disks, direct on-line information interchange, etc. If a creating agency sends records to NARA in an electronic media that NARA is not equipped to handle, these records are effectively lost.

5 NARA's Framework View

This section discusses the perspective and responsibilities of NARA in the provision of guidelines, policy, software for an interagency information interchange system, and an integrated data dictionary system for NARA-wide coordinated Data Administration and descriptive information management.

Since the archival process is a joint undertaking between NARA, the creating agencies, and the various agencies and researchers who retrieve information, much of the NARA Framework view has been discussed in the previous sections of this document. Accordingly, this section does not attempt to restate the information previously provided. Instead, this section concentrates on those aspects of the Framework view and policy that are unique to NARA.

NARA-wide coordinated Data Administration will depend on the automated support of a data dictionary system with an integrated user interface system. Such a system must be an implementation of the approved standard for data dictionary and should be designed to have features useful for information entry and retrieval within NARA, and features useful to querying by creating agencies and researchers.

5.1 Accessioning Policy for Formatted Records and Associated Descriptive Information

At the current time, the lack of consistency by outside agencies in following NARA prescribed procedures and formats in the submission of records to be archived is a problem that can be resolved without the loss of the submitted materials. Once the physical material is in the possession of NARA, it can always be reviewed by NARA personnel and appropriate descriptive material

41

can then be generated. As previously stated, this is not true with electronic records. Thus it is of primary importance that NARA begin now to develop the appropriate accessioning policy for how these electronic records are to be submitted, along with their descriptive information.

The "Recommendations for Document Transfer Standards and Their Integration into National Archives Policy," produced by NIST in conjunction with this document provides a discussion of the various document transfer standards that NARA should consider in the development of its final policy. Further, the document recommends the use of the Standard Page Description Language as the appropriate format for NARA to receive electronic documents.

The "Recommendations for Database and Data Dictionary Standards and Their Integration into National Archives Policy," also produced by NIST in conjunction with this document, provides standards recommendations in the areas of transferring and archiving electronic database information.

It must be recognized that no amount of regulations and policy statements will be successful unless NARA can supply outside agencies with some method of implementing them by way of automated software. Thus, NARA must establish the policy that, in order to achieve the needed interoperability in the exchange of electronic records, databases, and descriptive information, NARA will have to fund the production of some amount of computer software that will assist outside agencies in the implementation of NARA policy.

5.2 Preservation of Stored Records

The format and methods by which electronic records are stored are, in many ways, not as important as the retention of the information that describes these formats and methods. As long as this descriptive information is available, and the storage media of the subject electronic records is not unreadable by available devices, it is always possible to reformat and move electronic records to accommodate new advances in technology that make the old formats and methods obsolete.

For additional discussion on the standards that could be applied under NARA policy in this area, the reader should refer to the "Recommendations for Document Transfer Standards and Their Integration into National Archives Policy" and "Recommendations for Database and Data Dictionary Standards and Their Integration into National Archives Policy."

Since the formats and methods under which archived electronic records are retained will change through the years, it is vital that the preparation and retention of the descriptive information about these records and their formats be emphasized in NARA policy preparation. It cannot be emphasized enough that this descriptive information is the key that unlocks and makes useful any electronic records that are to be retained by the National Archives.

5.3 Media Assignment

At the current time there is no one electronic medium that would be appropriate for use in storage of all accessioned electronic records. Further, the central question in the issue of media assignment is really more one of the survivability of the information on the media and the availability of hardware to read the storage media. This is an area outside the scope of this document, and is therefore not addressed.

What is an appropriate area of discussion for this document is the need to continuously reevaluate NARA policy in this area. Due to future technological changes that NARA will have to accommodate, NARA policy in this area must ensure that storage media and storage formats do not fall too far behind the accepted usage of the time. This is necessary in order to avoid future problems with conversion of electronic records to new media. This is especially critical for the descriptive information about NARA's electronic records holdings.

Thus, NARA policy should be established to require regular, cyclical reviews at approximately five year intervals to determine the suitability of NARA electronic media and format procedures in view of current technology. This regular review must be done in order to ensure that NARA electronic storage and retrieval techniques do not become obsolescent. In developing this policy, NARA might wish to consider a two tier conversion approach in which the descriptive information about NARA electronic record holdings is transferred to a new storage medium more often then the actual electronic record holdings themselves. This approach would allow NARA to gain experience in the use of the new storage media or formats. Based on this experience, and in accordance with some evaluation criteria, NARA could then decide if it is appropriate to transfer all electronic record holdings to the new media.

5.4 Format and Media Conversion After Accession/Transfer

NARA policy covering this topic must be developed that accommodates two separate issues. The first issue is that discussed in the previous paragraph, the need to accomplish format and media conversion that may be required after accessioning in order to avoid future obsolescence of the storage media and to permit the distribution of records to current and future users with unknown software.

The second issue that NARA policy must consider is that, while records and descriptive information from other agencies should generally be received by NARA in prescribed standard formats and on, or via, the prescribed media, what will be done with those records received that are not in accordance with the prescribed standards. When the format of accessioned records do not meet NARA's policy guidelines for permanent storage there would be a cost associated with any format conversion required. A NARA policy must be developed that determines how, and by whom this cost is to be funded.

6 Future User Framework View

Future users of information in the National Archives should not change greatly from those who currently access information, although the number of users might increase. What will change will be the methods by which these users (creating agencies, researchers, and historians) choose to access information. As computer and network usage become increasingly integrated into our society, information access by electronic means will become the choice of many archives users.

6.1 Record Access and Retrieval Information

Unlike paper records that can be stored in small groups covering limited topics, electronic records are stored in vast amounts covering many, perhaps totally unrelated topics, on single storage devices. Without indexing information, an electronically stored record cannot be found. Thus record access and retrieval requires the availability of cross-referencing and indexing information. Availability of accurate cross-referencing and indexing information is especially critical to future users of the archives, since it is likely that no individuals will be available to the user who might remember where and how needed information was stored.

Another way in which electronic records are different from paper that is critical to future users is the issue of physical proximity of documents. Currently an archives user might be able to find many records of interest by physically reviewing other documents that are located in close proximity to the document that was initially retrieved. This is because paper documents that are related are often kept close together. Electronic storage devices do not necessarily keep information physically close. Instead the information is distributed on the storage device based on available space and device dependent requirements. This separation of information can increase when information is transferred to new storage media. The link that holds related information together is accurate cross-referencing and indexing information. An isolated individual record has little historical significance without cross-referencing other records sharing one or more similar characteristics.

Thus, to the future user, the most important aspect of the Framework is the development by NARA of appropriate cross-referencing and indexing information. Only if these types of descriptive information are captured and maintained by NARA will the accessibility of accessioned records by future users be assured.

6.2 Record Access Format and Media

It is well understood today that current electronic storage media does not have an infinite storage lifetime. Accordingly, it is common practice to periodically copy electronic records for preservation purposes, thus preventing possible loss of information due to media degradation. While periodic copying of electronic records within the same media is a solution to the problem of media degradation, it is not a solution to the format and media problems of future users. Even when a future user manages to locate a record, that record will not be usable, unless the record is stored in a format and media accessible to that future researcher. It is simply not possible to predict what format and media will be useful to future users of the archives. However, it can be said with certainty that, if NARA does not regularly transfer its electronic holdings to newer storage devices, the time will occur when electronic records will be lost. That is not to say they will be physically lost as one could lose a piece of paper. The loss will be the inability to read the record due to the unavailability of the devices that can read it. Thus, while the actual physical storage device holding the needed electronic record (i.e. magnetic disk) might be physically available, if the information cannot be read, it is the same as having shredded a piece of paper today. The piece of paper might be physically present, but the information it contained is lost forever.

6.3 Record Preservation

Based on the above discussion it should be clear that the issue of record preservation for future users of the archives is a twofold problem. First, the actual records must be on devices that make it physically available and, second, a record is not satisfactorily preserved for the future user unless it is cross-referenced and indexed in such a way as to make it identifiable.

Since, as time goes on, it will be necessary for NARA to change the physical storage devices utilized for electronic records, the key elements of the framework for future users are scheduling, descriptive, and retrieval information.

7 Data Administration Framework View

7.1 Centralized Coordination of Data Administration

The previous sections of this document have pointed out the importance of centralized coordination of Data Administration in order for NARA to control and make accessible electronic records information. This is not to say that there should only be one office in NARA for Data Administration. It is perhaps quite desirable to divide Data Administration in NARA into several different areas. What is critical is to have some part of the NARA organization that is identified as having the responsibility for centralized coordination of the various Data Administration activities occurring throughout NARA.

This centralized Data Administration office could ensure that the various Data Administration activities share standardized procedures, avoid duplication of effort, and share results on a regular basis.

The types of activities that would be coordinated under the centralized Data Administration function are those that analyze, define, manipulate, and maintain descriptive information about records, their origins, contents, storage format and media, accessibility, and use, etc. Descriptive information would include such things as record schedule information, life cycle tracking information, record access and retrieval information, cross-referencing information, and applications of this information.

7.2 Methodology and Automated Support for Data Administration

Under a centralized Data Administration function, a single coordinated Data Administration methodology could be defined and implemented as a part of NARA's policy for archiving of electronic records. This methodology would define the methods to be used throughout NARA in descriptive information analysis, definition, naming, storage, and access.

As part of this Data Administration methodology, it should be indicated that resources will be provided to put in place and maintain the needed automated support for NARA descriptive information. This automated support would be primarily supplied through a data dictionary system, also known as a repository. Further, as part of its policy NARA should require that this repository conform to the Information Resource Dictionary System (IRDS) standard. The IRDS standard was approved by the American National Standards Institute in October 1988 and was approved as Federal Information Processing Standard 156 (FIPS 156) during March 1989. As a FIPS standard, the IRDS will be required in future repository procurement in the Federal government.

If NARA chooses not to put in place a single coordinated Data Administration function with appropriate resources, then the lack of centralized methodology and automated support will lead to confusion, duplication of effort, and less use or misuse of descriptive information. The final result of this will be the eventual loss of electronic records information.

7.3 Descriptive Information Interchange

Due to the vital role played by descriptive information in the storage and retrieval of archived information, it is important that this information be maintained in a manner that can be utilized to satisfy information interchange requirements that are both internal and external to NARA. The internal information interchange needs of NARA would include the handling, exchanging, and retrieving of all descriptive information generated by the different functional areas of NARA in the process of accessioning electronic records. The external needs of creating agencies and researchers in locating, understanding, and gaining access to accessioned and transferred records must also be accommodated by this same system. By utilizing a repository based on the approved IRDS standard, the ability to exchange information is already accommodated.

8 Document Standards

For an in-depth discussion of the issue of document standards the reader should refer to Attachment C, "Recommendations for Document Transfer Standards and Their Integration into National Archives Policy." In summary, the recommendations of this document are as follows:

- Documents should normally be transferred to NARA in the format dictated by the Standard Page Description Language (SPDL). This is a non-revisable document format which can be transmitted, viewed, or printed by all conforming output devices. SPDL is the format that was utilized in the development of the Document Transfer System prototype software. While SPDL is not yet an approved standard, its development as a standard is high on the priority list of the NIST group involved in document transfer standards. In those cases where NARA has determined that it would be advantageous to utilize some other format, such differences must be clearly identified in the descriptive information repository system.

- This SPDL version of a document should be considered the final document.

- When appropriate, this SPDL version could be copied and converted to other standard formats for the purposes of keyword retrieval, but these copied versions are not to be considered the final document.

9 Form Standards

There currently exist a number of different standards that are utilized for the exchange of "forms" information in the business community. These are highly structured standards that are not suitable for the exchange of multiple types of electronic documents. These standards, such as the Electronic Data Interchange (EDI) series of Electronic Business Data Interchange (EBDI) Standards are intended for use in the exchange of such things as purchase orders and inventory control forms. Thus, these standards might be useful to NARA for the transfer and storage of completed government forms, but they would not be suitable for transfer or storage of textual type information.

10 Graphics Standards

At this time graphics standards can be broken down into two basic subsets. The first is the group of standards that are intended to standardize the functions of graphics programs. The existing standards in this area are as follows:

- Graphical Kernel System (GKS)
- Programmer's Hierarchical Interactive Graphics System (PHIGS)

Since the above standards are intended to standardize programming functions, or interfaces to programs, they would not be of any use to NARA in storing electronic records.

The second subset of graphics standards are those concerned with the transmission and storage of graphics information. These standards are:

- Computer Graphics Metafile for the Storage and Transfer of Picture Description Information (CGM)
- CCITT Group 3 or Group 4 Facsimile (FAX), this is a form of RASTER
- Initial Graphics Exchange Specification (IGES)

This set of graphics standards may be applicable for use by NARA, but only for those cases where NARA is maintaining files of strictly graphics information (e.g. engineering drawings, etc.).

For graphical type information within textual documents, the SPDL standard for documents will accommodate the graphical portion of the document. Accordingly, before any decision on use of a graphics standard could be made, NARA would have to survey the amount and type of graphical information to be archived.

For example, storing electronically generated document type information under the standards for RASTER graphics can be done, but would require much greater amounts of storage capacity then would be needed under other types of standards. Accordingly, while use of RASTER might be appropriate for electronically storing copies of older NARA holdings (produced by optical scanning devices), it would not be practical to use RASTER for currently produced textual information.

For a short description of each of the above standards see Attachment A, "Document Interchange Standards: Description and Status of Major Document and Graphics Standards," NISTIR 88-3851.

Another area of graphics standards in which NARA may have concerns is the area of digital cartographic data. While many organizations in the Federal Government have been maintaining digital cartographic data for some time, only recently has any concerted effort begun to standardize this information. Accordingly, if NARA is interested in this area, it might wish to establish a policy that requires creating agencies to be responsible for retaining custody of these records, at least until the standards in this area reach a higher level of maturity.

11 Database Standards

At the current time there is no one standard in the database area that could be utilized by NARA to store and retrieve all types of data from multiple types of databases. Most database transfer programs utilized in government and business are one time, database specific applications. These transfer programs are not standards, they usually rely on outside action to prepare the new database in the required format, and they are not intended, or designed, to serve as universal database translation and transfer programs.

The accessioning and storage of database records will require three distinct components until a more complete database standard is in place. The three components necessary to capture database information are: (1) the description of the original database schema in an Information Resource Dictionary (IRD) conformant with the IRDS standard; (2) one or more flat files of data from the original database, referenced by the IRD; and, (3) the description of the layout structure of the flat file(s), referenced by the IRD, and possibly stored in the same or another IRD.

For a more detailed discussion of database standards the reader should refer to "Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status."

12 Library Interchange Standard

The Machine Readable Cataloging (MARC) was developed and is supported by the Library of Congress for the exchange of Library of Congress cataloging information with libraries throughout the United States. The Society of American Archivists adopted the revised MARC format for manuscripts which is known as the MARC Archival and Manuscripts Control (MARC AMC). Accordingly, MARC AMC is now the de facto standard for exchange of lifecycle tracking information among archivists.

While the MARC AMC is a machine independent format for information exchange among archivists and librarians, it does not appear to be in direct use in any other area of the Federal government. Accordingly, it is doubtful that other Federal agencies would develop or purchase software that would utilize the MARC AMC just for the purpose of transmitting electronic documents to the National Archives. Further, as was stated in the NARA document titled The MARC Format and Life Cycle Tracking at the National Archives: A Study, the MARC AMC process control functions "while sufficient to meet the needs of smaller repositories, are not sophisticated enough to handle easily the complex and multiple needs of NARA." Thus, while the National Archives must maintain some level of compatibility with MARC AMC for use in the archivist and librarian communities, it is impractical to expect to utilize the MARC AMC format for data interchange to and from NARA by all other Federal agencies.

13 Generic Interchange Standards

There are a number of standards in existence that address the area of information interchange. This section addresses those that appear to be of possible use by NARA. For more information on these and other generic interchange standards the reader should refer to "Document Interchange Standards: Description and Status of Major Document and Graphics Standards" and "Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status." Both of these documents are provided as attachments to this document.

13.1 Abstract Syntax Notation One (ASN.1)

Using ASN.1, one can specify a formal description of data types and values. Such a description is called an Abstract Syntax.

This standard is described in ISO 8824. The Transfer Syntax is basically a bit level representation of the data being described by the Abstract Syntax. It is the Transfer Syntax that is sent between two Application Entities (i.e., users). The mapping from the data represented abstractly in ASN.1 to a Transfer Syntax is accomplished by using the Basic Encoding Rules (BER). The ASN.1 Basic Encoding Rules, which specify the derivation of a Transfer Syntax from an Abstract Syntax, are described in ISO 8825.

Due to its use in other standards that are recommended or required for inclusion in archives policy, it is also appropriate for ASN.1 to be used for other interchanges that fall outside of these specific standards.

13.2 Data Descriptive File (DDF)

Even if the National Archives were to focus exclusively on magnetic tape as its interchange medium, because of the future direction of technology the DDF standard would not be recommended for use in data transfer. Due to the current emphasis on the Government Open Systems Interconnection Protocol (GOSIP) standards for exchange of electronic records, ASN.1 should be considered as the basis of all interchange.

14 Additional Recommendation Information

Throughout this document there are a number of policy recommendations for NARA. This section provides some additional information concerning some of these recommendations. It is, of course, impossible for NARA to simultaneously implement the many and varied policy recommendations contained in this document. However, NARA could begin to implement some of the policy recommendations by establishing pilot projects that utilize elements of these policies, and their associated standards, to test their validity in "real world" applications. An example of a pilot project is the Document Transfer Prototype software developed as a part of this project. By performing this pilot project to develop a prototype, NARA has now tested the validity of this concept, and can now use this prototype as the basis for producing the requirements for a fully functional Document Transfer system.

14.1 Responsibility of NARA to Define Interchange Policy

In order to successfully move toward the combined paper and electronic document environment, NARA policy should define and make recommendations for the use of specific software standards for record and descriptive information interchange, in both the near and long term.

NARA policy should define a strategy on conformance testing for standards use validation on both database and electronic document records submitted to NARA. A procedure must be established for identifying if records received comply with the format, media, and presentation software standards defined in NARA policy. Included as aspects of this policy must be:

- The requirement for parser(s) that check record and descriptive information files to determine whether the files are encoded according to NARA-defined standards Policy

- The requirement that all non-standard records be converted to NARA standard format in some manner; otherwise those non-conforming records will be unmaintainable, and eventually inaccessible, as software and hardware evolve

- A procedure, or set of procedures, for responding to creating agencies that send records in formats, or media that do not comply with NARA Policy; in this case, NARA options are:

    - to inform the creating agency of the non-conformance, and then send records back to the creating agency for conversion (by creating agency) according to NARA defined standards

    - to inform the creating agency of the non-conformance, give the creating agency a certain deadline by which the agency must achieve conformance, and, in the interim, have NARA personnel convert records according to NARA defined standards

    - to inform creating agency of the non-conformance, and, if the assumption is that the creating agency will never comply with NARA Policy guideline, have NARA personnel continually convert records according to NARA defined standards

Due to the limited nature of standards in the database area, current NARA policy should direct continued attention to the database transfer area. One action NARA could take now would be to survey and determine the types of databases that are expected to be accessioned in the foreseeable future. From such a survey, it is possible that a near-term NARA electronic interchange policy could be defined with a limited scope of the database transfer capability that must be supported at this time.

14.2 Responsibility of NARA to Define Data Administration Policy

14.2.1 Assessment of NARA Data Administration Position

The current Life Cycle Management function has done an excellent job of tracking descriptive information on record handling within NARA. Also, several other Data Administration activities performed at NARA have been quite successful within their current scope of activity, but taken as a group they are not directly coordinated.

14.2.2 Need for Coordination of Data Administration Activities

As previously stated in the document, NARA policy should specify coordination of all Data Administration activities within the organization. This would provide unified direction, to ensure that all necessary activities are undertaken by all area specific Data Administration offices. This would reduce unintentional contradictions and duplication of effort. It would improve accessioned and transferred record accessibility and retrieval both by NARA personnel and by other users. Further it would serve to put in place the organizational structure needed to provide researchers and creating agencies with access to cross-referencing descriptive information that will permit the identification of records with some common aspects.

14.2.3 Need for Data Administration Support for Record Storage and Access

In order to gain access to any particular record or records, it is necessary to know the format, media, and software configuration of the record or records before it, or they, **can** be read. Further, additional information is necessary in order to physically locate files of a particular type or category that may be stored in a variety of media, formats, and in different locations.

Accordingly, initial NARA policy in this area should assign a centralized Data Administration activity to define, integrate, automate, and maintain descriptive information on record contents, storage, retrieval, and access for all database and electronic documents held by NARA. In the future, NARA may also wish to place under this centralized Data Administration function, the responsibility to coordinate the maintenance of the location and descriptive information on all NARA holdings.

14.2.4 Need for Descriptive Information Automation

NARA policy establishing a centralized coordinated Data Administration function should specify that planning be undertaken to put in place the needed automation for a Data Administration information system. Such an information system must allow different offices within NARA to maintain and access a shared database of descriptive information on accessioned records.

The NARA information automation policy should also require the eventual automation of all information control forms, such as the planned record schedule automation. This automation should be directed to combine and integrate all such forms with the proposed automated Data Administration information system.

14.3 Responsibility of NARA to Define Policy for Systems Integration

NARA policy should require planning for integration of all automated systems so that the various systems at NARA can interchange data, use shared formats defined by Data Administration activities, and eventually work interactively. Without such a policy for systems integration, NARA will be spending funds for duplicate efforts, contradictory methods, and for developing translation software to communicate between individual systems.

Finally, if NARA's policy and actions do not now begin to address the problem of electronic records management, NARA faces the situation of being overrun with non-retrievable electronic records that cannot be recovered once the information on "how to access" the records is lost or inaccessible. Irretrievable electronic records are of no use either to the creating agencies or to NARA, and will simply "be there" uselessly taking up space.

NISTIR 88-3851

Document Interchange Standards: Description and Status of Major Document and Graphics Standards

Judi Moline

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
(Formerly National Bureau of Standards
National Computer and Telecommunications Laboratory
Office Systems Engineering Group
Gaithersburg, MD 20899

September 1988

NISTIR 88-3851

Document Interchange Standards: Description and Status of Major Document and Graphics Standards

Judi Moline

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
(Formerly National Bureau of Standards)
National Computer and Telecommunications Laboratory
Office Systems Engineering Group
Gaithersburg, MD 20899

September 1988

National Bureau of Standards became the
National Institute of Standards and Technology
on August 23, 1988, when the Omnibus Trade and
Competitiveness Act was signed. NIST retains
all NBS functions. Its new programs will encourage
improved use of technology by U.S. industry.

U.S. DEPARTMENT OF COMMERCE
C. William Verity, Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Ernest Ambler, Director

DOCUMENT INTERCHANGE STANDARDS:
Description and Status of Major Document and Graphics Standards

ABSTRACT

Document interchange standards have emerged in response to two distinct needs. First, there is the need to interchange documents among workstations and tools in the office environment. Second, there is the need to exchange versions of a document between an author and a publisher. This document describes standards which attempt to satisfy those needs. Each relevant standard is presented in summary form and includes the following information: name of standard, number of standard, status, scope, description, use, and references.

KEY WORDS:

CGI, CGM, DFR, DIF, document interchange, DSSSL, EDI, FAX, GGCA, GKS, IGES, ODA, ODIF, ODL, PHIGS, Raster, SDIF, SGML, SPDL, standards, TRIF.

1. Introduction

Paper documents as we know them today are complex with their rich variety of contents: assorted modes of emphases, photographs, diagrams, tables, maps, etc. Electronic documents have the potential of all the parts we have today plus many new ones: animated sequences, aural components, dynamic charts, tables, and graphs, etc. Paper documents are easily interchanged but not easily updated and revised. Electronic documents have the potential for easy updating and revision. Further, they can be easily interchanged among machines if appropriate steps are taken. However, if this potential is to be realized, standards must be used so that the document can be readily transferred from machine to machine thus allowing updating and revision independent of the originating machine.

Although the standards process is not new, with the current transition to electronic offices, standards are being developed rapidly. It is difficult to keep up to date unless one is in the center of the activity. However, more and more of us need a general understanding of what is available as we work to determine the offices and the systems of today and the future.

It is to fulfill this perceived need that this document originated. It will not answer all the detailed questions about the standards. It will, however, provide a general introduction to the current document and graphics standards.

1.1 Document Overview

The following information is provided for each standard: name of standard, reference number, status, scope, description, use, and references.

1.2 Acknowledgments

I would like to thank Lawrence A. Welsch and David K. Jefferson for encouraging me to publish these materials which have been very helpful for me and others who have used them.

2. *Document Standards Summaries*

The standards covered include the following:

Abbreviation	Name of Standard	Doc. No.
CGI	Computer Graphics Interfacing Techniques for Dialogues with Graphical Devices	ISO DP 9636:1988
CGM	Computer Graphics Metafile for the Storage and Transfer of Picture Description Information	ISO 8632:1986 ANSI X3.122-1986 FIPS PUB 128
DFR	Document Filing and Retrieval	ISO/IEC JTC1/SC 18/WG 4 N1264 ISO/IEC JTC1/SC 18/WG 4 N1265
DIF	Document Interchange Format	NBSIR 84-2836
DSSSL	Document Style Semantics and Specification Language	ISO/IEC JTC1/SC 18/WG 8 N606
EDI	Electronic Data Interchange (series of Electronic Business Data Interchange (EBDI) Standards)	
	Data Element Dictionary	ANSI X12.3:1986
	Application Control Structure	ANSI X12.6:1986
	Functional Acknowledgment	ANSI X12.20:1986
	Data Segment Dictionary	ANSI X12.22:1986
FAX	CCITT Group 4 Facsimile	Recommendations T.5 and T.6
GGCA	Geometric Graphics Content Architecture	ISO 8613-8:1988
GKS	Graphical Kernel System	ISO 7942:1985 ANSI X3.124-1985 FIPS PUB 120
IGES	Initial Graphics Exchange Specification	NBSIR 86-3359 (IGES 3.0)
	Digital Representation for Communication of Product Definition Data (Based on IGES 3.0.)	ASME/ANSI Y14.26M-1987
	Initial Graphics Exchange Specification	NBSIR 88-3813 (IGES 4.0)
ODA	Office Document Architecture	ISO 8613:1988
ODIF	Office Document Interchange Format	ISO 8613-5:1988
ODL	Office Document Language	ISO 8613-5:1988
PHIGS	Programmer's Hierarchical Interactive Graphics System	ISO DIS 9592:1988 ANSI X3.144-1988
Raster	Raster Graphics Content Architecture	ISO 8613-7:1988
SDIF	SGML Document Interchange Format	ISO DIS 9069:1987
SGML	Standard Generalized Markup Language	ISO 8879:1986 FIPS PUB 152
SPDL	Standard Page Description Language	ISO/IEC JTC1/SC 18/WG 8 N561
TRIF	Tiling Raster Interchange Format	Proposed Extension to ISO 8613-7:1988

Other acronyms found in the document include:

ANSI	American National Standards Institute
ASME	American Society of Mechanical Engineers
DIS	Draft International Standard
DP	Draft Proposal
FIPS PUB	Federal Information Processing Standards Publication
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
JTC	Joint Technical Committee 1 of ISO and IEC
NBS	National Bureau of Standards (now NIST)
NBSIR	National Bureau of Standards Internal Report
NIST	National Institute of Standards and Technology
SC	Subcommittee
TTG	Tiling Task Group
WG	Working Group

NAME OF STANDARD: Information Processing Systems - Computer Graphics - Interfacing Techniques for Dialogues with Graphical Devices (CGI)

NUMBER OF STANDARD: ISO DP 9636:1988

STATUS: International Draft Proposal (1988)

SCOPE: "This standard establishes the conceptual model, functionality, minimum conformance requirements, and encodings of the Computer Graphics Interface (CGI). This standard defines a usable set of CGI functions that is expected to satisfy the following needs of a majority of the computer graphics community:

1. Provide an interface standard for computer graphics software package implementors.
2. Provide an interface standard for computer graphics device manufacturers and suppliers.
3. Provide an inquiry and response mechanism for graphics device capabilities, characteristics, and states.
4. Provide a standard graphics escape mechanism to access non-standard graphics device capabilities.
5. Provide for future functional extension of the CGI.

"In addition to the CGI functionality, device classes and constituency profiles are defined. The device classes included in the CGI are output (OUT), input (IN), and input/output (OUTIN). Constituency profiles allow subsets of the CGI functions and features to be defined to suit particular constituencies of users. There is also provision for constituency profiles to be registered after the standard is published." (ISO/DP9636 "1.1 Scope")

DESCRIPTION: "The Computer Graphics Interface (CGI) is a standard functional and syntactical specification of the control and data exchange between device-independent graphics software and one or more Virtual Devices.

"The syntax of the CGI is presented in this standard in a binding-independent and encoding-independent specification. Any similarity of the examples or function specifications to a particular language or encoding technique is purely accidental, unless explicitly stated otherwise.

"The functions specified provide for the representation of a wide range of two-dimensional pictures and control over their display on a wide range of graphical devices. The functions are split into groups that perform device and CGI session control, which specify the data representations used, control the display of the pictures, perform basic drawing actions, control the attributes of the basic drawing actions, and provide access to non-standard device capabilities.

"Part 1 of this standard gives an overview of this multipart standard, explains the relationship between the parts and the relations to other standards, describes the graphics standards reference model used, and defines certain constituency policies. Parts 2 to 6 specify the CGI functions for different functional areas using an abstract syntax.

"Additional separate standards will define standard data-stream encodings, procedural library bindings, and single-entry-point procedural bindings of the CGI." (ISO/DP9636 "1.2 Field of Application")

USE: "The CGI standard describes graphical services provided by a Virtual Graphics Device. The model for description of these services is expressed in terms of graphical capabilities of a single instance of a hypothetical graphics device. In all but the simplest of computing environments, CGI functions alone will not be sufficient to provide complete control over a device. Additional functions, not included in this standard will likely be needed. Examples of such functions include:

— means to configure (sets of) physical devices to be accessed as CGI Virtual Devices.

— means to control a device capable of offering CGI-defined services as well as other, non-CGI-defined services, such as those implied by ISO 2022 and ISO 6429.

— means to differentiate among separate instances of CGI Virtual Devices in the same computing environment; and

— means of defining or determining communication paths from CGI clients to CGI Virtual Devices.

"In some cases, other standards exist that describe the functions required. For example, various communications standards address the needs of the last point above. In other cases no standards may exist, but the tasks indicated are outside the scope of a Computer Graphics Standard. The second point mentioned above is such an example." (ISO/DP9636 "1.2 Relationship of CGI to a Computing Environment")

REFERENCES:

Arnold, D.B. and P.R. Bono. *CGM and CGI: Metafile and Interface Standards for Computer Graphics.* Berlin: Springer-Verlag, 1988.

ISO/DP9636 (CGI) - Interim Draft *Information Processing Systems - Computer Graphics - Interfacing Techniques for Dialogues with Graphical Devices - Functional Specification* April 15, 1988.

Powers, Thomas, Andrea Frankel, and David Arnold. "The Computer Graphics Virtual Device Interface." *IEEE Computer Graphics and Applications* (August 1986) 6:8, 33-41.

NAME OF STANDARD: Computer Graphics Metafile for the Storage and Transfer of Picture Description Information (CGM)

NUMBER OF STANDARD: ISO 8632:1986 and ANSI X3.122-1986

STATUS: International Standard (1986) and American National Standard (1986)

SCOPE: "The Computer Graphics Metafile provides a file format suitable for the storage and retrieval of picture description information. The file format consists of an ordered set of elements that can be used to describe pictures in a way that is compatible between systems of different architectures and devices of differing capabilities and design.

"The elements specified provide for the representation of a wide range of pictures on a wide range of graphical devices. The elements are split into groups that delimit major structures (metafiles and pictures), that specify the representations used within the metafile, that control the display of the picture, that perform basic drawing actions, that control the attributes of the basic drawing actions and that provide access to non-standard device capabilities.

"The Metafile is defined in such a way that, in addition to sequential access to the whole metafile, random access to individual pictures is well-defined; whether this is available in any system that uses this Standard depends on the medium, the encoding and the implementation.

"In addition to a Functional Specification, three standard encodings of the metafile syntax are specified. These encodings address the needs of applications that require minimum metafile size, minimum effort to generate and interpret, and maximum flexibility for a human reader or editor of the metafile." (ISO 8632-1:1986 1 "Scope and Field of Application")

DESCRIPTION: "The Computer Graphics Metafile provides a file format suitable for the storage and retrieval of picture information. The file format consists of a set of elements that can be used to describe pictures in a way that is compatible between systems of different architectures and devices of differing capabilities and design." (ISO 8632)1:1986 0.1 Purpose)

"The main reasons for producing a standard computer graphics metafile are:
a) to allow picture information to be stored in an organized way on a graphical software system;
b) to facilitate transfer of picture information between different graphical software systems;
c) to enable picture information to be transferred between graphical devices;
d) to enable picture information to be transferred between different computer graphics installations."

The parts of the international standard are as follows:

1. Functional Specifications
2. Character and Coding
3. Binding and Coding
4. Clear Text Encoding

USE: "The use of this standard is strongly recommended when one or more of the following situations exist:

— A graphics metafile is maintained at a central facility for a decentralized system that employs graphics devices of different makes and models that must utilize the data.

— A graphics metafile is required to preserve picture data when conversion or migration from one graphics system to another is necessary and the two systems are not necessarily compatible.

— A graphics metafile is intended for information interchange between a source system and a target system that are not necessarily compatible."
(FIPS PUB 128 - Computer Graphics Metafile "9. Applicability")

REFERENCES:

ANSI X3.122-1986 *American National Standard for Information Systems - Computer Graphics - Metafile for the Storage and Transfer of Picture Description Information*.

Arnold, D.B. and P.R. Bono. *CGM and CGI: Metafile and Interface Standards for Computer Graphics*. Berlin: Springer-Verlag, 1988.

Henderson, Lofton, Margaret Journey, and Chris Osland. "The Computer Graphics Metafile." *IEEE Computer Graphics and Applications*. (August 1986) 6:8, 24-32.

ISO 8632-1986 *Information Processing Systems - Computer Graphics Metafile for the Storage and Transfer of Picture Description Information*.

FIPS PUB 128 *Computer Graphics Metafile (CGM)* 1987 March 16 adopts ANSI X3.122-1986.

NAME OF STANDARD: Information Processing - Document Filing and Retrieval (DFR)

NUMBER OF STANDARD: None

STATUS: Proposed DP (1988)

SCOPE: "This Standard is for a Distributed Application located in the Application Layer of the Reference Model for Open Systems Interconnection (see ISO 7498).

"It should be noted that a DFR will provide storage for an open-ended set of document types; the content of the documents is transparent to the DFR.

"This Standard serves the following important fields of application:

— capability for large capacity document storage for use by multiple users in a distributed system;

— ordered filing and multi-key retrieval of documents;

— structured organization of groups of documents;

— storage of an open-ended number of different document types (i.e. the content of a document is transparent to the DFR);

— filing and referencing documents outside of the document storage (i.e. a non-electronic hard copy document);

— adjoining attributes to a document independent from the document's content;

— capabilities to store, retrieve and delete documents of the document store whatever their content;

— capabilities to search for, order, retrieve, and delete single documents or groups of documents using document attributes;

— protection against unauthorized storage and retrieval of documents;

— capabilities to control concurrent access to the documents."
(ISO/IEC JTC1/SC 18/WG 4 N1264 "1. Scope and Field of Application")

DESCRIPTION: "The Document Filing and Retrieval Application forms part of a series of standards defining the application needed in the area of office automation, as it is described in the Distributed Office Application Model. This standard provides the functionality which directly supports the user in an office environment. Thus, Document Filing and Retrieval is not aiming to a general standardization of all types of filestores, as they may exist in computing systems. Rather, it concentrates on the filing and retrieval of documents, as it is a task of office work. Document Filing and Retrieval only aims to standardize the model of such documentstores and the services and protocols defining the principles, how clients can access such documentstore servers, whereas clients and servers reside on different nodes of a distributed office system." (ISO/IEC JTC1/SC 18/WG 4 N1264 "0. Introduction")

USE: "The Document Filing and Retrieval Application provides the capability for large capacity non-volatile document storage to multiple users in a distributed office system. This facility is particularly useful in an environment where a large population of desktop workstations that have limited storage capacity require access to large expensive storage devices." (ISO/IEC JTC1/SC 18/WG 4 N1264 "0. Introduction")

"This standard will deal with individual document filing and retrieval servers. The Standard governs the interactions of a document filing and retrieval client and a single document filing and retrieval server. Future standardization will consider the facilities of a distributed filing and retrieval server system and the need for inter-server protocols. It is intended that the results of the initial standardization work be extensible and support this latter work." (ISO/IEC JTC1/SC 18i/WG 4 N1264 "Note".)

REFERENCES:

ISO/IEC JTC1/SC 18/WG 4 N1264 *Information Processing - Document Filing and Retrieval (DFR) - Part 1: Abstract Service Definition and Procedures*

ISO/IEC JTC1/SC 18/WG 4 N1265 *Information Processing - Document Filing and Retrieval (DFR) - Part 2: Protocol Specification*

NAME OF STANDARD: (Navy) Document Interchange Format (DIF)

NUMBER OF STANDARD: NBSIR 84-2836

STATUS: Navy Standard (1984)

SCOPE: At the time that DIF was developed (before 1984), there was no standard which addressed "even a majority of the control information required by text processing systems."

"Based on ANSI standards X3.98, X3.64, and ISO 6429, 13 functions were found to be already standardized. It was decided to use the framework provided by ANSI standard X3.64 to encode the remaining 29 functions." (NBSIR 84-2836 pages 2 and 3.)

DESCRIPTION: "The forty-four control functions defined by DIF have been divided into six classes according to their primary purpose. The six classes are Break functions, Document Format functions, Page Format functions, Line Format functions, Rendition functions, and Miscellaneous functions." (NBSIR 84-2836 page 5.)

USE: DIF provides transfer of basic text and formatting instructions.

REFERENCES:

NBSIR 84-2836 *Document Interchange Format* April 1984.

NAME OF STANDARD: Information Processing - Text and Office Systems - Document Style Semantics and Specification Language (DSSSL)

NUMBER OF STANDARD: None

STATUS: Proposed DP (1988)

SCOPE: "Document Style Semantics and Specification Language (DSSSL) primarily provides the means of specifying the desired appearance of a composed SGML document, independent of any formatting process and any specific type of formatter. As such it:

a. is used to specify the relationships between the SGML logical elements as expressed in the source DTD and the intended result;

b. defines formatting and style semantics and a corresponding machine-processable syntax to describe the typographic style and layout of a document; and

c. allows the typographic semantics to be extended as required by specific applications.

"This International Standard is intended to be extensible in order to accommodate future developments in formatting and other document processing technologies." (ISO/IEC JTC1/SC18/WG 8 N606 "1.1 Scope")

DESCRIPTION: "The DSSSL semantics are designed to be interpretable by a wide variety of page layout, formatting, and other document processing systems, including printing, display, and data base loading and extraction. The specification language portion of DSSSL has two purposes:

a. To connect DSSSL constructs to SGML constructs.

b. To express relationships among DSSSL constructs.

"The specification language is, therefore, designed to accommodate two categories of language component, namely datatypes and operators." (ISO/IEC JTC1/SC18/WG 8 N606 "4.1 Introduction")

USE: "This International Standard is intended for use in a wide variety of SGML application environments, including:

a. Electronic publishing (e.g., production publishing, workgroup publishing, desktop publishing, database publishing, etc.)

b. Electronic galley creation, pagination, and imposition

c. Printing specifications (offset lithography, letterpress, gravure, demand printing)

d. Display and/or on-line information

"This International Standard provides a means to specify document processing characteristics of a formatted document which may be represented in SPDL (Standard Page Description Language) or in some other format.

"Documents that exist solely in final formatted form are not within the field of application of this International Standard." (ISO/IEC JTC1/SC18/WG 8 N606 "1.2 Field of Application")
REFERENCES:

ISO/IEC JTC1/SC18/WG 8 N606 *Information Processing - Text and Office Systems - Document Style Semantics and Specification Language (DSSSL)* July 1988

NAME OF STANDARD: Electronic Data Interchange (EDI) series of Electronic Business Data Interchange (EBDI) Standards

NUMBER OF STANDARD:	ANSI X12.3:1986 Data Element Dictionary
	ANSI X12.6:1986 Application Control Structure
	ANSI X12.20:1986 Functional Acknowledgment
	ANSI X12.22:1986 Data Segment Directory

STATUS: American National Standards (1986)

SCOPE: "The ANSI X12 Data Interchange Standards consist of transaction set standards, a data dictionary, a segment directory, and transmission control standards." (ANSI *An Introduction to Electronic Data Interchange* July 1987, p. 2.)

ANSI X12.3:1986 - "This standard provides the specifications of the data elements that comprise the transaction sets described in the X12 series." (ANSI X12.3:1986 "1. Purpose and Scope")

ANSI X12.6:1986 - "This American National Standard is used in the definition of the structure and content of business transaction for computer-to-computer interchange. This standard describes the control segments used to envelop loops of data segments, to envelop transaction sets, and to envelop groups of related transaction sets. This standard does not define any specific transaction or group of transactions to be standardized. The actual data segments of this standard are diagrammed in American National Standard for Electronic Business Data Interchange - Data Segment Directory, ANSI X12.22-1986; the description of those control segments is contained in this standard." (ANSI X12.3:1986 "1. Purpose and Scope")

ANSI X12.20:1986 - "The purpose of this standard is to define the control structures for a set of acknowledgments to indicate the results of the syntactical analysis of the electronically encoded documents. The encoded documents are the transaction sets, which are grouped in functional groups, used in defining transactions for business data interchange. This standard does not cover the semantic meaning of the information encoded in the transaction sets." (ANSI X12.3:1986 "1. Purpose and Scope")

ANSI X12.22:1986 - "This standard provides the definitions and formats of the data segments used in the construction of the X12 series." (ANSI X12.22:1986 "1. Purpose and Scope")

DESCRIPTION: "Transaction set standards define the procedural format and data content requirements for specified business transactions, such as purchase orders.

"The data dictionary defines the precise content for data elements used in building transaction sets.

"The segment directory provides the definitions and formats of the data segments used in building transaction sets.

"The transmission control standards define the formats for the information required to interchange data. These controls are already in use by some industry groups." (ANSI *An Introduction to Electronic Data Interchange* July 1987, p. 2.)

USE: EDI is used to standardize the format and content of data to be interchanged between two computers. Subsets of the total EDI Standards package will be selected based on the needs of the potential interchange partners.

REFERENCES:

ANSI *An Introduction to Electronic Data Interchange* July 1987.

ANSI X12.3:1986 *American National Standard for Electronic Business Data Interchange - Data Element Dictionary.*

ANSI X12.6:1986 *American National Standard for Electronic Business Data Interchange - Application Control Structure.*

ANSI X12.20:1986 *American National Standard for Electronic Business Data Interchange - Functional Acknowledgment.*

ANSI X12.22:1986 *American National Standard for Electronic Business Data Interchange - Data Segment Directory.*

NAME OF STANDARD: CCITT Group 4 Facsimile

NUMBER OF STANDARD: CCITT Recommendations T.5 and T.6

STATUS: International Standard (1984)

SCOPE: The Group 4 Facsimile Standard has two parts. Recommendation T.5 "defines the general aspects of Group 4 facsimile apparatus. The Group 4 facsimile coding scheme and facsimile control functions are defined in Recommendation T.6." (CCITT *Red Book* "Recommendation T.5 2 Scope")

DESCRIPTION: "The Group 4 apparatus provides the means for direct document transmission from any subscriber to any other subscriber.

"All apparatus participating in the international Group 4 facsimile service has to be compatible with each other at the basic level defined in this Recommendation. Additional operational functions may be invoked.

"The range of data rates is described in Section 6. Detailed arrangements on a national level are left to the Administrations concerned, as it is recognized that national implementation of the Group 4 facsimile service on various types of networks may involve national operation at different data throughput rates.

"The page is the basis for facsimile message formatting and transmission. Both A4 and North American paper formats are taken into account.

"Facsimile coding schemes are applied in order to reduce the redundant information in facsimile signals prior to transmission.

"The apparatus must have the ability to reproduce facsimile messages. The content, layout and format of facsimile messages must be identical at the transmitting and receiving apparatus.

"The reproducible area is defined within which facsimile messages are assured to be reproduced.

"The Group 4 facsimile apparatus should provide means for automatic reception. In addition Class II/III apparatus should provide means for automatic reception of Teletex and mixed mode documents.

"All Classes of Group 4 facsimile apparatus shall incorporate the functions defined as basic for the Group 4 facsimile service in Section 3.2 below. In addition, optional functions can be incorporated. In this Recommendation, the optional functions are divided into CCITT standardized options and nationally and/or privately specified options." (CCITT *Red Book* "Recommendation T.5 3.1 Basic Characteristics")

"Facsimile coding schemes consist of the basic facsimile coding scheme and optional facsimile coding schemes.

"Facsimile coding schemes are specified assuming that transmission errors are corrected by control procedures at a lower level.

"The basic facsimile coding scheme is the two-dimensional coding scheme which is in principle the same as the two-dimensional coding scheme of Group 3 facsimile specified in Recommendation T.4.

"Optional facsimile coding schemes are specified not only for black and white images but also for grey scale images and colour images.

"Facsimile coding control functions are used in facsimile user information in order to change facsimile parameters or to invoke the end of facsimile block." (CCITT *Red Book*, "Recommendation T.6 1.2.1 Facsimile coding schemes and coding control functions")

USE: "Group 4 facsimile is used mainly on public data networks (PDN) including circuitswitched, packet-switched, and the integrated services digital network (ISDN). The apparatus may also be used on the public switched telephone network (PSTN) where an appropriate modulation process will be utilized.

"The procedures used with Group 4 facsimile apparatus enable it to transmit and reproduce image coded information essentially without transmission errors.

"Group 4 facsimile apparatus has the means for reducing the redundant information in facsimile signals prior to transmission.

"The basic image type of the Group 4 facsimile apparatus is black and white. Other image types, e.g. grey scale image or colour image, are for further study.

"There are three classes of Group 4 facsimile terminals:

- Class I - Minimum requirement is a terminal able to send and receive documents containing facsimile encoded information.
- Class II - Minimum requirement is a terminal able to transmit documents which are facsimile encoded. In addition, the terminal must be capable of receiving documents which are facsimile coded, Teletex coded, and also mixed-mode documents.
- Class III - Minimum requirement is a terminal which is capable of generating, transmitting and receiving facsimile coded documents, Teletex coded documents, and mixed-mode documents." (CCITT *Red Book* "Recommendation T.5 1 General")

"Group 4 facsimile apparatus shall be capable of handling:

a. the basic end-to-end control procedures as defined in Recommendation T.62;
b. document interchange protocol as defined in Recommendation T.73;
c. the basic facsimile coding scheme as defined in Recommendation T.6;
d. the control functions associated with the basic facsimile coding scheme as defined in Recommendation T.6.

"All classes of Group 4 apparatus shall have the following provisions for facsimile messages:

a. provision for scanning the documents to be transmitted;
b. provision for receiving and presenting hard or soft copies of the documents.

"In addition Group 4 Class II apparatus shall have provision for receiving and displaying basic Teletex and mixed mode documents.

"In addition to the requirements for Group 4 Class II apparatus, Class III apparatus shall have provisions for generating and transmitting basic Teletex and mixed mode documents.

"Basic page formatting functions are as follows:

a. vertical page orientation;
b. paper size of ISO A4;
c. reproducible area/printable area is defined taking into account ISO A4 and North American paper formats and ISO standard 3535." (CCITT *Red Book* "Recommendation T.5 3.2 Basic Functions")

REFERENCES:

CCITT *Red Book Vol VII fascicle VII.3 Terminal Equipment and Protocols for Telematic Services*. "Recommendation T.5 General Aspects of Group 4 Facsimile Apparatus." Geneva 1985.

CCITT *Red Book Vol VII fascicle VII.3 Terminal Equipment and Protocols for Telematic Services*. "Recommendation T.6 Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus." Geneva 1985.

NAME OF STANDARD: Geometric Graphics Content Architectures (GGCA)

NUMBER OF STANDARD: ISO 8613-8:1988

STATUS: International Standard (1988)

SCOPE: "The purpose of this International Standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, geometric graphics elements and raster graphics elements, all potentially within one document.

"ISO 8613 applies to the interchange of documents by means of data communication or the exchange of storage media." (ISO 8613-8 "1 Scope")

DESCRIPTION: "This part of ISO 8613:

- defines a geometric graphics content architecture that can be used in conjunction with the document architecture defined in ISO 8613-2;
- defines an interface which allows the use of content structured [sic] according to ISO 8632 within documents structured according to ISO 8613-2;
- defines those aspects of positioning and imaging applicable to the presentation of this geometric graphics content architecture in a basic layout object;
- defines the presentation attributes applicable to this geometric graphics content architecture;
- describes a content layout process, which together with the document layout process described in ISO 8613-2, describes the layout of geometric graphics content in basic layout objects and determines the dimensions of these basic layout objects."

(ISO 8613-8 "1.3")

USE: "ISO 8613-8 applies to documents that are structured according to the architecture defined in ISO 8613-2 that include geometric graphics content, consisting of a descriptive representation of picture description information as an ordered set of elements such as lines, arcs, polygons, attributes for these drawing elements, elements that structure the content portion, etc. using the Computer Graphics Metafile (CGM) and its binary encoding defined in ISO 8632-1 and ISO 8632-3, respectively." (ISO 8613-1 "6.7 Part 8 Geometric graphics content architectures")

REFERENCES:

ISO 8613-1:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format - Part 1 - Introduction and General Principles.*

ISO 8613-8:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format - Part 8 - Geometric Graphics Content Architectures (GGCA).*

NAME OF STANDARD: Computer Graphics - Graphical Kernel System (GKS) Functional Description

NUMBER OF STANDARD: ISO 7942:1985 and ANSI X3.124-1985

STATUS: International Standard (1985) and American National Standard (1985)

SCOPE: "This American National Standard specifies a set of functions for computer graphics programming, the Graphical Kernel System (GKS). GKS is a basic graphics system for applications that produce computer generated two dimensional pictures on line graphics or raster graphics output devices. It supports operator input and interaction by supplying basic functions for graphical input and picture segmentation. It allows storage and dynamic modification of pictures. A fundamental concept in GKS is the workstation, consisting potentially of a number of input devices and a single output device. Several workstations can be used simultaneously. The application program is allowed to adapt its behavior at a workstation to make best use of workstation capabilities. This standard includes functions for storage on and retrieval from an external graphics file. Last, but not least, the functions are organized in upward compatible levels with increasing capabilities.

"For certain parameters of the functions, GKS defines value ranges as being reserved for registration or future standardization. The meanings of these values will be defined using the procedures established in an International Standard under development (Procedures for registration of graphical items)." (ANSI X3.124:1985 "1 Scope and Field of Application")

DESCRIPTION: "The Graphical Kernel System (GKS) provides a functional interface between an application program and a configuration of graphical input and output devices. The functional interface contains all basic functions for interactive and non-interactive graphics on a wide variety of graphics equipment.

"The interface is at such a level of abstraction that hardware peculiarities are shielded from the application program. As a result a simplified interface presenting uniform output primitives, and uniform input classes is obtained.

"A central concept both for structuring GKS and for realizing device independence is introduced, called the workstation.

"The facilities for picture manipulation and change are introduced via the segment facilities, the dynamic attributes and the transformations.

"The concept of multiple workstations allows simultaneous output to and input from various display systems. Facilities for internal and external storage are provided by special workstations together with the possibility of transferring graphical entities directly from the special workstation for internal storage to other workstations.

"Not every GKS implementation needs to support the full set of functions. Twelve levels are defined to meet the different requirements of graphics systems. Each GKS implementation provides at least the functions of one level. The levels are upward compatible." (ANSI X3.124:1985 GKS "4.2 Introduction")

USE: "The Graphical Kernel System (GKS) is a set of basic functions for computer graphics programming usable by many graphics producing applications." Use of "this standard

1. allows graphics application programs to be easily transported between installations,

2. aids graphics applications programmers in understanding and using graphics methods, and
3. guides device manufacturers on useful graphics capabilities."

(ANSI X3.124-195 GKS "Abstract")

REFERENCES:

ANSI X3.124-1985 *American National Standard for Information Systems - Computer Graphics - Graphical Kernel System (GKS) Functional Description*

Bono, Peter R., Jose L. Encarnacao, F. Robert A. Hopgood, and Paul J.W. ten Hagen. "GKS - The First Graphics Standard." *IEEE Computer Graphics and Applications*. (July 1982) 2:5, 9-23.

FIPS PUB 120 Graphical Kernel System adopts ANSI X3.124-1985.

NAME OF STANDARD: Initial Graphics Exchange Specification (IGES)

NUMBER OF STANDARD: ASME/ANSI Y14.26M-1987 Digital Representation for Communication of Product Definition Data (Based on Version 3.0 of the Initial Graphics Exchange Specification published as NBSIR 86-3359.)

STATUS: American National Standard (1987)

SCOPE: "This document establishes information structures to be used for the digital representation and communication of product definition data. Use of the specification established herein permits the compatible exchange of product definition data used by various (CAD/CAM) Computer Aided Design and Computer Aided Manufacturing) systems." (ASME/ANSI Y14.26M-1987 "1.1 Purpose")

DESCRIPTION: "This specification defines a file structure format, a language format, and the representation of geometric, topological, and non-geometric product definition data in these formats. Product definition data represented in these formats will be exchanged through a variety of physical media. The specific features and protocols for the communications media are the subject of other standards. The methodology for representing product definition data in this specification is extensible and independent of the modeling methods used.

Chapter 1 is general in nature and defines the overall purpose and objectives of this specification. "Chapter 2 defines the communications file structure and format. It explains the function of each of the sections of a file. The geometry data representation in Chapter 3 deals with two- and three-dimensional edge-vertex models and with simple surface representations. Chapter 4 specifies non-geometric representations, including common drafting practices, data organization methods, and data definition methods.

"In Chapters 3 and 4, the product is described in terms of geometric and non-geometric information, with non-geometric information being divided into annotation, definition, and organization. The geometry category consists of elements such as points, curves, and surfaces that model the product. The annotation category consists of those elements which are used to clarify or enhance the geometry, including dimensions, drafting notation, and text. The definition category provides the ability to define specific properties or characteristics of individual or collections of data entities. The organization category identifies groupings of elements from geometric, annotation, or property data which are to be evaluated and manipulated as single items." (ASME/ANSI Y14.26M-1987 "1.2 Field of Application")

USE: IGES is used "to describe and communicate the essential engineering characteristics of physical objects as manufactured products. Such products are described in terms of their physical shape, dimensions, and information which further describes or explains the product. The processes which generate or utilize the product definition data typically include design, engineering analysis, production planning, fabrication, material handling, assembly, inspection, marketing, and field service." (ASME/ANSI Y14.26M-1987 "1.4 Concepts of Product Definition")

REFERENCE

ASME/ANSI Y14.26M-1987 *Digital Representation for Communication of Product Definition Data.*

H. Grabowski and R. Glatz, "IGES Model Comparison System: A Tool for Testing and Validating IGES Processors," *IEEE Computer Graphics and Applications*, (November 1987) 47-57.

IGES *Technical Illustrations Application Guide*, April 1987.

IGES *Recommended Practices Guide*, November 1987.

IGES *Electrical Application Guide*, March 1987.

MIL-D-28000 *Digital Representation for Communication of Product Data: IGES Application Subsets*, 22 December 1987.

MIL-STD-1840A *Automated Interchange of Technical Information*, 22 December 1987.

NBSIR 86-3359 *Initial Graphics Exchange Specification, Version 3.0.*

NBSIR 88-3813 *Initial Graphics Exchange Specification, Version 4.0.*

NAME OF STANDARD: Office Document Architecture (ODA) and Interchange Format (ODIF)

NUMBER OF STANDARD: ISO 8613:1988

STATUS: International Standard (1988)

SCOPE: "The purpose of this international standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used in the documents may include graphic characters, geometric graphics elements, and raster graphics elements, all potentially within one document.

"NOTE : ISO 8613 is designed to allow for extensions, including typographical features, colour, spreadsheets and additional types of content such as sound." (ISO 8613-1:1988 (E) "1.1)

DESCRIPTION: ODA was developed to allow the interchange of documents from one word processor to another. Page layout is handled according to some precise semantics which strive to be content independent. The page or sets of pages are specified denoting margins, columns, character path, line progression, etc., which detail the placement of rectangular "blocks," with content, specifically characters, image, and graphics to be poured in to occupy various areas on the page. (Adler)

The parts of the standard are as follows:

1. General Introduction
2. Document Structures
4. Document Profile
5. Office Document Interchange Format (ODIF) (see ODIF, ODL, and SDIF)
6. Character Graphics Content Architectures
7. Raster Graphics Content Architectures (see Raster and TRIF)
8. Geometric Graphics Content Architectures (see GGCA)

USE: ODA/ODIF is specifically designed for the interchange and replication of office documents in exact format. The design strives to be content-independent in order to allow for future content architectures such as audio information or possible mathematical and scientific equations.

REFERENCES:

Adler, Sharon C. "SGML and ODA: Two Standards for the Interchange of Documents," *<TAG> The SGML Newsletter,* Volume 1, Issue 4, 1-3.

Horak, Wolfgang. "Office Document Architecture and Office Document Interchange Formats: Current Status of International Standardization" *Computer* (October 1985), 50-60.

ISO 8613:1988 (E) *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format.*

NAME OF STANDARD: Office Document Interchange Format (ODIF)

NUMBER OF STANDARD: ISO 8613-5:1988

STATUS: International Standard (1988)

SCOPE: "The purpose of this International Standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, geometric graphics elements and raster graphics elements, all potentially within one document.

"ISO 8613 applies to the interchange of documents by means of data communication or the exchange of storage media." (ISO 8613-5 "1 Scope")

DESCRIPTION: "This part of ISO 8613:

— defines the format of the data stream used to interchange documents structured in accordance with ISO 8613-2;

— defines the representation of the constituents which may appear in an interchanged document."

(ISO 8613-5 "1.3")

"ODIF is an abstract data syntax in which the constituents and attributes of the document are represented by a hierarchy of data structures and data items, specified using the abstract syntax notation ASN.1 defined in ISO 8824.

"The coded representation of each data structure or data item is obtained by applying a set of encoding rules." (ISO 8613-5 "4.1 ODIF")

"The ODIF data stream is described in terms of a set of data structures, called 'interchange data element', which represent the constituents (document profile, object descriptions, object class descriptions, presentation styles, layout styles and content portion descriptions) of a document. The formats of the interchanged data element according to ODIF are defined using the Abstract Syntax Notation One (ASN.1) specified in ISO 8824." (ISO 8613-1 "6.4 Part 5 Office document interchange format (ODIF))"

USE: A document structured in accordance with ISO 8613 may be represented for interchange by the Office Document Interchange Format (ODIF). Since ODIF is a data structure specified using ASN.1, it is intended for use in an OSI environment. (ISO 8613-5 "4 Document representations")

REFERENCES:

ISO 8613-1:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format*

ISO 8613-5:1988 *Information Processing - Text and Office Systems - Office Document Interchange Format (ODIF)*

NAME OF STANDARD: Office Document Language (ODL)

NUMBER OF STANDARD: ISO 8613-5:1988

STATUS: International Standard (1988)

SCOPE: "The purpose of this International Standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, geometric graphics elements and raster graphics elements, B all potentially within one document.

"ISO 8613 applies to the interchange of documents by means of data communication or the exchange of storage media." (ISO 8613-5 "1 Scope")

DESCRIPTION: "This part of ISO 8613:

— defines the format of the data stream used to interchange documents structured in accordance with ISO 8613-2;

— defines the representation of the constituents which may appear in an interchanged document."

(ISO 8613-5 "1.3")

"ODL uses the Standard Generalized Markup Language (SGML) specified in ISO 8879. It consists of a standard set of SGML names and markup conventions for representing the constituents and attributes of a document." (ISO 8613-1 "6.4 Part 5 Office document interchange format (ODIF)")

USE: A document structured in accordance with ISO 8613 may be represented for interchange by the Office Document Language (ODL) in conjunction with the SGML Document Interchange Format (SDIF). "ODL is particularly appropriate for systems that share information through marked-up text files, especially where human users can access the markup directly." (ISO 8613-5 "4 Document representations")

REFERENCES:

ISO 8613-1:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format*

ISO 8613-5:1988 *Information Processing - Text and Office Systems - Office Document Interchange Format (ODIF)*

NAME OF STANDARD: Computer Graphics - Programmer's Hierarchical Interactive Graphics System (PHIGS)

NUMBER OF STANDARD: ISO DIS 9592:1988 and ANSI X3.144:1988

STATUS: Draft International Standard (1988) and American National Standard (1988)

SCOPE: "This American National Standard specifies a set of functions for computer graphics programming, the Programmer's Hierarchical Interactive Graphics System (PHIGS). PHIGS is a graphics system for applications that produce computer generated pictures on line graphics or raster graphics output devices. It supports operator input and interactions by supplying basic functions for graphical input and hierarchical picture definition. It allows for storage, and dynamic modification of pictures.

"A fundamental concept in PHIGS is the *workstation*, consisting of a number of input devices and a single output device. Several workstations can be used simultaneously. The application program is allowed to adapt its behavior at a workstation to make best use of workstation capabilities. A second fundamental concept is the *centralized structure store*, where graphical information is stored and edited.

"This American National Standard includes functions for storage on and retrieval from an external graphics file." (ANSI X3.144:1988 "1 Scope and Field of Application")

DESCRIPTION: "The Programmer's Hierarchical Interactive Graphics System (PHIGS) provides a functional interface between an application program and a configuration of graphical input and output devices. The functional interface contains basic functions for dynamic interactive hierarchical graphics on a wide variety of graphics equipment.

"The interface is at such a level of abstraction that hardware peculiarities are shielded from the application program.

"PHIGS defines only a language independent nucleus of a graphics system. For integration into a language, PHIGS is embedded in a language dependent layer containing the language conventions, for example, parameter and name assignment." (ANSI X3.144:1988 "4.2 Overview")

USE: "This Standard:

— allows graphics application programs to be easily transported between installations.

— aids graphics applications programmers in understanding and using graphics methods.

— guides device manufacturers on useful graphics capabilities.

— performs many functions currently performed by graphics applications; thus, offloading the graphics application development effort.

"This Standard defines an application level programming interface to a hierarchical interactive and dynamic graphics system. Hence it contains functions for:

— outputting graphical primitives.

— controlling the appearance of graphical primitives with attributes.

— controlling graphical workstations.

— controlling 2D & 3D transformations and coordinate systems.

— generating, modifying, and controlling groups of primitives called structures.

— modifying the hierarchical relationship of structures.

— obtaining graphical input.

— archiving and retrieving structures and structure hierarchies.

— inquiring the capabilities and states of the graphics system.

— handling errors."
(ANSI X3.144:1988 "Abstract")

REFERENCES:

ANSI X3.144:1988 *American National Standard for Information Systems - Computer Graphics - Programmer's Hierarchical Interactive Graphics System (PHIGS) Functional Description*

ISO DIS 9592:1988 *Programmer's Hierarchical Interactive Graphics System.*

Shuey, David, David Bailey, and Thomas P. Morrissey. "PHIGS: A Standard, Dynamic, Interactive Graphics Interface." *IEEE Computer Graphics and Applications* (August 1986) 6:8, 50-57.

NAME OF STANDARD: Raster Graphics Content Architecture

NUMBER OF STANDARD: ISO 8613-7:1988

STATUS: International Standard (1988)

SCOPE: "The purpose of this International Standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, geometric graphics elements and raster graphics elements, all potentially within one document.

"ISO 8613 applies to the interchange of documents by means of data communications or the exchange of storage media." (ISO 8613-7 "1 Scope")

DESCRIPTION: "This part of ISO 8613 defines:

— the raster graphics content architectures that can be used in conjunction with the document architecture defined in ISO 8613-2;

— the internal structure of content portions that are structured according to a raster graphics content architecture;

— those aspects of positioning and imaging applicable to the presentation of raster graphics contents in a basic layout object;

— a content layout process which, together with the document layout process defined in ISO 8613-2, specifies the method for determining the dimensions of basic layout objects for raster graphics content portions;

— the presentation and content portion attributes applicable to raster graphics content architectures."
(ISO 8613-7 "1.3")

USE: "ISO 8613-7 applies to documents that are structured according to the architecture defined in ISO 8613-2 that include raster graphics content, consisting of a descriptive representation of pictorial information provided by an array of picture elements (pels), encoded according to facsimile or bitmap encoding." (ISO 8613-1 "6.6 Part 7 Raster graphics content architectures")

REFERENCES:

ISO 8613-1:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format - Part 1 - Introduction and General Principles.*

ISO 8613-7:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format - Part 7 - Raster Graphics Content Architectures.*

NAME OF STANDARD: Information Processing - SGML Support Facilities - SGML Document Interchange Format (SDIF)

NUMBER OF STANDARD: ISO DIS 9069:1987

STATUS: Draft International Standard (1987)

SCOPE: "This International Standard specifies a data structure known as the SGML Document Interchange Format (SDIF). SDIF enables a document conforming to ISO 8879, which might be stored in several entities, to be packed into a data stream for interchange in a manner that will permit the recipient to reconstitute the separate entities.

"SDIF also allows related documents to be included in the data stream, such as covering letters, transmittal forms, catalog cards, formatting procedures, font resources, or the 'document profile' required by a document architecture." (ISO DIS 9069 "1 Scope")

DESCRIPTION: "The SDIF data stream represents one or more SGML document entities, and zero or more SGML subdocument, SGML text, and data entities, as defined in ISO 8879." (ISO DIS 9069 "5 Description of the Data Stream")

USE: "The SGML Document Interchange Format shall be used solely for the interchange of SGML documents, as defined in ISO 8879, among SGML systems.

"Interchange can be by means of data communications in Open Systems Interconnection or other environments, or by the exchange of storage media." (ISO DIS 9069 "2 Field of Application")

REFERENCES:

ISO 8613-1:1988 *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format*

ISO 8613-5:1988 *Information Processing - Text and Office Systems - Office Document Interchange Format (ODIF)*

ISO DIS 9069:1987 *Information Processing - SGML Support Facilities - SGML Document Interchange Format (SDIF)*

NAME OF STANDARD: Standard Generalized Markup Language (SGML)

NUMBER OF STANDARD: ISO 8879:1986

STATUS: International Standard (1986)

SCOPE: "This International Standard specifies an abstract syntax known as the Standard Generalized Markup Language (SGML). The language expresses the description of a document's structure and other attributes, as well as other information that makes the markup interpretable.

"This International Standard specifies a reference concrete syntax that binds the abstract syntax to specific characters and numeric values, and criteria for defining variant concrete syntaxes.

"This International Standard defines conforming documents in terms of their use of components of the language.

"This International Standard defines conforming systems in terms of their ability to process conforming documents and to recognize markup errors in them.

"Specifies how data not defined by this International Standard (such as images, graphics, or formatted text) can be included in a conforming document." (ISO 8879 "1 Scope")

DESCRIPTION: SGML was designed to interchange documents without regard to how the information is formatted. This allows for the use of the information in many different formats. SGML was designed to be application independent, and as such can be used in conjunction with a database application. The user is allowed to interact with and to modify the logical structures which are a primary part of his application. An SGML document may be processed by any formatter (for a formatting application) which has been suitably enabled with an SGML parser and other entity-management software. The SGML notation may be used to describe both logical and layout structures, if the format of the document is also to be interchanged. A set of standardized formatting semantics are to be provided by DSSSL. (Adler 2-3)

USE: SGML is specifically designed for the world of publishing and the management and control of the information which may take form in many types of documents.

"SGML can be used for publishing in its broadest definition, ranging from single medium conventional publishing to multi-media data base publishing. SGML can also be used in office document processing when the benefits of human readability and interchange with publishing systems are required." (ISO 8879 "0 Introduction")

REFERENCES:

Adler, Sharon C. "SGML and ODA: Two Standards for the Interchange of Documents," *<TAG> The SGML Newsletter*, Volume 1, Issue 4, 1-3.

ISO 8879:1986 *Information Processing - Text and Office Systems - Standard Generalized Markup Language (SGML)*, First Edition - 1986-10-15.

ISO 8879:1986(E) *Technical Errata* as of April 30, 1987.

Smith, Joan M. *The Standard Generalized Markup Language(SGML): Guidelines for Authors.* British National Bibliography Research Fund Report 27. Great Britain: The British Library, 1987.

NAME OF STANDARD: Standard Page Description Language (SPDL)

NUMBER OF STANDARD: None

STATUS: Proposed DP (1988)

SCOPE: "The scope of this International Standard is the specification of a device-independent and process-independent description of images of documents in fully composed, non-revisable form. Such documents may utilize the full capabilities of imaging devices which may include high-resolution printing machinery and softcopy output devices.

"This International Standard is intended to be extensible in order to accommodate future developments in imaging technology.

"This International Standard is intended to be used in a variety of configurations meeting a variety of connectivity needs. It is specifically compatible with use over OSI networks.

"In addition to specifying how document images are represented, this International Standard specifies how additional information called *printing instructions* affects the document image. Printing instructions may be supplied with the request to print the document by means of a *print access protocol*." (ISO 3rd Working Draft "1.1 Scope")

DESCRIPTION: "The Standard Page Description Language is capable of representing all content types for fully composed, non-revisable documents. Any combination of the following types of content can be represented; any content may in [sic] black-and-white, gray-scale, or full colour; and content types may be intermixed in any way in the same document.

— character

— raster graphics

— geometric graphics."
(ISO 3rd Working Draft "1.1 Scope")

USE: "This International Standard is intended for use in a wide variety of application environments, including:

— electronic publishing (including production publishing, workgroup publishing, desktop publishing, database publishing, electronic prepress, etc.)

— office systems

— information networks

— demand printing

"This International Standard provides a straightforward and efficient method of representing documents which are generated by ODA systems to presentation devices. It also provides a capability for similarly representing documents generated by SGML systems whose formatting is described by DSSSL.

"This International Standard allows for document presentation to be disjoint in both time and place from the document creation and formatting processes. It is specifically intended that SPDL document descriptions will be:

— sent directly to presentation systems which are accessed via a local connection

— sent to proximate or remote presentation systems via OSI or non-OSI networks, and

— stored or interchanged for the purpose of presentation at other times or at other locations." (ISO 3rd Working Draft "1.2 Field of Application")

REFERENCES:

ISO/IEC JTC1/SC 18/WG 8 N561 *Information Processing - Text and Office Systems - Standard Page Description Language (SPDL)*, 3rd Working Draft - 1988-02-19.

NAME OF STANDARD: Tiled Raster Interchange Format (TRIF)

NUMBER OF STANDARD: Proposed Extension to 8613-7:1988

STATUS: Proposed to ANSI X3V1 (1988)

SCOPE: "The purpose of this International Standard is to facilitate the interchange of documents.

"In the context of ISO 8613, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, geometric graphics elements and raster graphics elements, all potentially within one document.

"ISO 8613 applies to the interchange of documents by means of data communications or the exchange of storage media." (ISO 8613-7 "1 Scope")

DESCRIPTION: "This part of ISO 8613 defines:

a. the tiled raster graphics content architectures that can be used in conjunction with the document architecture defined in ISO 8613-2;

b. the internal structure of content portions that are structured according to a tiled raster graphics contents in a basic layout object;

c. those aspects of positioning and imaging applicable to the presentation of tiled raster graphics contents in a basic layout object;

d. a content layout process which together with the document layout process defined in ISO 8613-2, specifies the method for determining the dimensions of basic layout objects for tiled raster graphics content portions;

e. the presentation and content portion attributes applicable to tiled raster graphics content architectures."

(TTG/88-14 TRIF 2.0 Tiled Raster Graphics Content Architecture"1.3")

USE: "The tiling scheme developed provides a format that supports operation on a subset of an image without requiring other portions of the image to be accessed. For large format documents this provides a way to interchange images between systems of various capabilities."

Further, the "tile format was developed for interchange that could also reasonably be used for storage and retrieval without necessarily requiring translation."

The following restrictions for use were made to ease user implementation:

— "This interchange format deals only with bi-tonal (black and white) data. Pixels are assumed to be square.

— "A tile is a rectangular region in a page in which all regions have the same dimensions (are regular) and no part of any region overlaps any other region. They are positioned in a fixed grid, determined by partitioning the page into tile-sized areas.

— "For the purposes of this interchange format, the application profile restricts all tiles to being square. Square tiles have the desirable attribute of being easily rotated. Tiles are allowed to be absent...

- "A single tile size is desirable to limit the burden on implementors of the interchange standard. The tile size is specifically 512 by 512 pels...
- "Only one page (one single raster image) is allowed per document.
- "Any given tile is to be encoded as T.6 compressed data, as bitmap data, or is specified as all foreground or all background..."

(TTG/88-20 Preliminary User Requirements for Tiled Raster Graphics TRIF 2.0)

REFERENCES:

TTG/88-14 *TRIF 2.0 Tiled Raster Graphics Content Architecture,* proposed to ANSI X3V1 by the Tiling Task Group, February 1988.

TTG/88-20 *Preliminary User Requirements for Tiled Raster Graphics TRIF 2.0,* 11 March 1988.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** *(See instructions)*	1. PUBLICATION OR REPORT NO. NISTIR 88-3851	2. Performing Organ. Report No.	3. Publication Date September 1988

4. TITLE AND SUBTITLE

Document Interchange Standards: Description and Status of Major Document and Graphic Standards

5. AUTHOR(S)

Judi Moline

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
(FORMERLY NATIONAL BUREAU OF STANDARDS)
U. S. DEPARTMENT OF COMMERCE
GAITHERSBURG, MD 20899

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

Same as above

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

Document interchange standards have emerged in response to two distinct needs. First, there is the need to interchange documents among workstations and tools in the office environment. Second, there is the need to exchange versions of a document between an author and a publisher. This document describes standards which attempt to satisfy those needs. Each relevant standard is presented in summary form and includes the following information: standard name, standard number, status, scope, description, use, and references.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

CGI/CGM, DFR, DIF, document interchange, EDI, FAX, GGCA, GKS, IGES, ODA/ODIF/ODL/Raster/TRIF, PHIGS, SGML/SDIF/DSSSL, SPDL, standards

13. AVAILABILITY

☐ Unlimited
☒ For Official Distribution. Do Not Release to NTIS
☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

15. Price

Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status

Attachment B:

Framework for the Exchange and Preservation of Electronic Records

Prepared for the

National Archives and Records Administration

By the

Information Systems Engineering Division
National Institute of Standards and Technology

Written by
Margaret H. Law

With Assistance from
Leonard Gallagher
and Tim Boland

January, 1989

Database, Data Dictionary, Interchange, and User Interface Standards: Description and Status

TABLE OF CONTENTS

Abstract Syntax Notation One (ASN.1) 1

Data Descriptive File (DDF) 4

File Transfer, Access, and Management (FTAM) 7

Information Resource Dictionary System (IRDS) 11

CCITT Data Communications Networks Message Handling System (MHS) 15

Database Language NDL 19

Remote Database Access (RDA) 22

Database Language SQL 24

ISO Information Processing Systems Open System Interconnection Series of Virtual Terminal (VT) Standards . 28

STANDARD NAME: Abstract Syntax Notation One (ASN.1)

STANDARD NUMBER:

ISO 8824:1987	Abstract Syntax Notation One
ISO 8825:1987	Encoding Rules for ASN.1
ISO 8824/Add1	ASN.1 Extensions
ISO 8825/Add1	Extension Encoding Rules

STATUS:

Abstract Syntax Notation One (ASN.1) and its associated Encoding Rules were adopted in 1987 by the International Organization for Standardization (ISO). It is expected that in 1989, ISO and the International Electrotechnical Commission (IEC) will adopt the recently specified ASN.1 Extensions for additional data types, subtypes, and other features. The X3 technical committee for this project in the United States plans to submit the combined package for adoption as an ANSI standard as soon as possible.

There are no known commercial "stand-alone" implementations of the ASN.1 standards, but every vendor of an Open Systems Interconnection (OSI) conforming product must have the capability to read and write objects defined by ASN.1. Such products have already appeared in the marketplace with front-end and back-end ASN.1 processors.

SCOPE:

ASN.1 is a language for defining data types and values in the OSI protocols, and is particularly used for Application Layer and Presentation Layer protocols. ASN.1 is used for defining documents in the Office Document Architecture (ODA) and Interchange Format (ODIF) family of standards and for data interchange in the Information Resource Dictionary System (IRDS) and Remote Data Access (RDA) standards. It is sufficiently general to be useful for defining arbitrary and complex data types and value structures.

The ASN.1 Encoding Rules specify how ASN.1 defined structures are represented as strings of 8-bit bytes. The Encoding Rules are used for interchanging occurrences of ASN.1 structures among heterogeneous environments.

DESCRIPTION:

ASN.1 specifies syntax rules for defining abstract structures. It consists of a kernel of universal data types along with various constructor types that allow specification of highly complex or nested structures. The primitive types include boolean, integer, bitstring, octetstring, and various character string types, with real and enumerated types and subtype definitions added in the forthcoming extension. Constructor types include sequence, set, choice, and tagged. All constructor types can be nested to any depth. Other useful types include an object-identifier type for referencing ASN.1 defined objects in other standards, generalized time, universal time, and explicit reference to eight different character sets.

The ASN.1 standard specifies four different classes of data types: Universal, Application, Private, and Context-specific. Universal types are defined only by the ASN standards; Application types are defined in terms of other standards; Private class is for user specifications; and Context-specific is meaningful only in the context of a higher level data type definition.

With ASN.1 one can specify a formal description of data types and values; such a description is called an Abstract Syntax. This standard is described in ISO 8824. The Transfer Syntax is basically a bit level representation of the data being described by the Abstract Syntax. It is the Transfer Syntax that is sent between two Application Entities (i.e., users). The mapping from the data represented abstractly in ASN.1 to a Transfer Syntax is accomplished by using the Basic Encoding Rules (BER). The ASN.1 Basic Encoding Rules, which specify the derivation of a Transfer Syntax from an Abstract Syntax, are described in ISO 8825.

The ASN.1 Encoding Rules specify explicitly how an occurrence of an ASN.1 defined structure is represented linearly as a sequence of 8-bit bytes. Each data type in an abstract type definition is represented as an ordered triple: type, length, value. The "type" identifies a Universal type or is a reference to some previously defined application, private, or context-specific type; the "value" is the actual representation; and the "length" is the number of 8-bit bytes needed to represent the value. The Encoding Rules specify the representation of each Universal data type as a sequence of 8-bit bytes. Application of the rules for linearizing the constructor types thereby produces a linear representation for each occurrence of any object defined by ASN.1.

GENERAL USE:

ASN.1 and its Encoding rules are appropriate for the interchange of arbitrary objects among heterogeneous environments. If the ASN.1 definition is known at both the "source" and the "target" of any interchange, then the representation produced at the source can be understood by the target. Such ASN.1 definitions could, themselves, be interchanged as "definition" data types.

USE IN ACCESSIONING ELECTRONIC RECORDS:

ASN.1 is the representation mechanism used in the Office Document Architecture (ODA) and Interchange Format (ODIF) standards, and in both the Information Resource Dictionary System (IRDS) standard and the Remote Data Access (RDA) standard for data interchange.

ASN.1 is expected to become the representation mechanism for all data management standards, including SQL. It is now clear that ASN.1 will be in widespread use in standards that will be part of the NARA policy. Accordingly it is of vital importance that NARA remain aware of other applications of ASN.1 that fall outside of the specific standards noted here.

REFERENCES:

ISO 8824, Information Processing Systems - OSI- Specification of Abstract Syntax Notation One (ASN.1), ANSI, 1430 Broadway, New York, May 1987.

ISO 8825, Information Processing Systems - OSI - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1), ANSI, 1430 Broadway, New York, May 1987.

ISO/IEC Draft Addendum 1, ASN.1 Extensions, June 1988.

STANDARD NAME: Data Descriptive File (DDF)

STANDARD NUMBER: ANSI/ISO 8211:1985 and FIPS PUB 123, Data Descriptive File

STATUS:

The specification for a Data Descriptive File (DDF) for Information Interchange was adopted in 1985 by the International Organization for Standardization (ISO) and in 1986 by the American National Standards Institute (ANSI). The ANSI standard was then adopted as a Federal Information Processing Standard (FIPS) in September of 1986, published as FIPS PUB 123.

A very small number of commercial "stand-alone" implementations of the DDF standard are available, since most implementations are written by users themselves. Some large user organizations and standards organizations are using the DDF specification for application specific, structured file interchange. In particular, the Proposed Standard for Digital Cartographic Data (see references), issued by the National Committee for Digital Cartographic Data Standards, specifies usage of the DDF for representing and transferring its data elements.

SCOPE:

The DDF specifies syntax for describing and representing a file of records. It serves much the same purpose as the Abstract Syntax Notation One (ASN.1) standard but is more focused on the characteristics of a classical COBOL or FORTRAN file consisting of fixed or variable length records composed of named fields of data. Even though the DDF predates ASN.1, it suffers from a lack of recognition in Open Systems Interconnection (OSI) environments where, for example, the OSI File Transfer protocol (FTAM) is defined exclusively in terms of ASN.1. Although specified independently of any transfer media or communications style, the DDF is more at home in the magnetic tape interchange environment from which it emerged. Due to the focus on OSI environments, DDF may no longer be the most appropriate interchange medium even for a magnetic tape environment.

DESCRIPTION:

The DDF standard specifies a Data Descriptive Record (DDR), which describes the characteristics of each data field, and a sequence of Data Records (DR's), which contain the actual data occurrences. The DDR is analogous to an ASN.1 definition and the DR's are analogous to an ASN.1 encoding. The DDR and DR records have the same structure, consisting of "leader", "directory", and "data" portions.

The "leader" is a sequence of 24 characters that gives the total record length in characters, codes for the level and type of the record, and information for reading the "directory" portion. The "directory" establishes integer "tags" that correspond to data fields in the "data" portion of each record and gives starting positions and lengths for all such fields.

The "data" portion of the DDR contains a "data descriptive field" for each of the "user data fields". Each data descriptive field specifies a data type and associates a data name or a reserved word with each tag. The "data" portion of each DR consists of data fields containing the raw data to be interchanged. Each data field in a DR is an instance of the user data structure and data type defined by the DDR data descriptive field with the corresponding field tag. Data Names in the DDR correspond to data values in a DR if and only if they have identical tags.

The standard provides for three implementation levels, depending on the complexity of the data files to be interchanged. Level 1 supports multiple fields containing simple, unstructured character strings. Level 2 supports Level 1 and also processes multiple fields containing structured data comprising a variety of data types. Simple data types include: character strings, numeric values from ISO 6093 (ANSI X3.42), and bit strings; whereas, structured data types include vectors, multi-dimensional arrays, and mixed type structures. Level 3 supports Level 2 and also allows hierarchical data structures. The best example of hierarchical data supported by Level 3 is a nested hierarchy of repeating groups from a COBOL record.

The standard specifies a special "short-hand" notation for interchange files that consist only of fixed-length records with data fields in which the DR's have identical leader and directory values. In this situation, the "leader" and "directory" of the first DR apply to all subsequent DR's and the subsequent "leader" and "directory" portions can be omitted.

GENERAL USE:

The Data Descriptive File is used for the interchange of files of records among heterogeneous environments. It is particularly used for interchanging COBOL or FORTRAN files of records on magnetic tape. Other structures can also be interchanged provided they are first represented as "records" or "files of records". DDF may no longer be the best choice for interchanging textual documents, however. Due to recent focus on Open Systems Interconnection (OSI) standards for exchange of magnetic records, DDF has become less popular and the ASN.1 presentation layer protocol is now recognized as the interchange standard of choice.

USE IN ACCESSIONING ELECTRONIC RECORDS:

If the National Archives focuses exclusively on magnetic tape as its interchange medium, then the DDF could be used to describe structured files of data records. However, because of the current market direction, even in a magnetic tape environment, it may not be the best choice for interchanging textual documents.

If the National Archives focuses exclusively on Open Systems Interconnection standards for exchange of electronic records, which is the current direction of most market products, then the DDF is much less appropriate and ASN.1 should be considered as the basis of all interchange.

REFERENCES:

ANSI/ISO 8211:1985, American National Standard for Information Systems - Specification for a Data Descriptive File for Information Interchange, ANSI, 1430 Broadway, New York, NY 10018, approved February 28, 1986.

Federal Information Processing Standard (FIPS) Publication (PUB) 123, Specification for a Data Descriptive File for Information Interchange (DDF), September 1986.

ACSM, "Proposed Standard for Digital Cartographic Data", The American Cartographer, Journal of American Congress on Surveying and Mapping, Vol 15, No 1, January 1988.

STANDARD NAME: File Transfer, Access, and Management (FTAM)

STANDARD NUMBER: ISO IS 8571

Part 1 - General Introduction
Part 2 - Virtual Filestore Definition
Part 3 - File Service Definition
Part 4 - File Protocol Specification

STATUS:

ISO approved FTAM as an International Standard in 1988, as part of the Open Systems Interconnection (OSI) set of standards. The three addenda to the FTAM Standard have not yet been approved. These three addenda include: the Filestore Management addendum that is expected to be approved in late 1989; the Overlapped Access addendum that is expected to be approved in early 1990; and the Protocol Conformance addendum that is expected to be approved in late 1990.

The Implementors' Agreements based on this FTAM Standard are referred to as the FTAM Phase 2 Agreements. Released in 1987, these Phase 2 Agreements provide facilities for full file transfer and record-level access. Three different FTAM document types and four different National Institute of Standards and Technology (NIST) document types are defined. This Implementors' Agreement provides for all services defined in the FTAM standard, except for restart, recovery, and concurrency.

The planned FTAM Phase 3 Agreements, which will be retroactively compatible with the FTAM Phase 2 Agreements, will provide restart, recovery, and concurrency capabilities, and enlarge on the set of document types currently defined. The FTAM Phase 3 Implementors' Agreement document is expected to be completed by early 1989.

SCOPE:

The services that FTAM provides to the user are the facilities to: (1) communicate about files without specific knowledge of the other system; (2) express explicitly what the user requires for file transfer, access, or management; (3) specify uniform file properties; (4) specify record-level file access and positional file transfer; and, (5) define detailed file management.

The FTAM Standard General Description deals with the basic FTAM terminology and broad concepts. The File Service Definition provides an overview of FTAM services provided to the user. The Virtual Filestore section gives information on the central model used by FTAM. Finally, the File Protocol specification describes in detail the protocol interactions necessary to accomplish the FTAM activity.

DESCRIPTION:

FTAM is a two-party file transfer protocol, in which an initiator of the file activity controls and directs the action, and a responder reacts to the initiator in a passive role. An FTAM implementation may act as initiator, responder, or both. Three-party file transfer is a subject of discussion for the future.

The FTAM service can be described as a series of regimes, where a regime is an environment which may be entered and exited via confirmed services. The first or outermost regime is the application association regime; this involves setting up an FTAM activity within the context of an association.

FTAM is defined in terms of Functional Units and Service Classes. Service Classes are described in terms of Functional Units; some Functional Units are mandatory within a Service Class and some Functional Units are optional. The Functional Units in FTAM are Kernel, Limited File Management, Enhanced File Management, Read, Write, Grouping, Recovery, and Restart.

As a Functional Unit, the Kernel is the basic set of FTAM capabilities. Limited File Management deals with the ability to create, delete, and interrogate properties of files. Enhanced File Management deals with the ability to change file properties. Grouping supports the joining or concatenations of likely FTAM requests for efficiency purposes.

File attributes are globally unique and may be seen by anybody accessing the file. Activity attributes are particular to a connection and are only visible to the user of the connection. Using FTAM, a user may query the values of these attributes and possibly change these values.

FTAM embodies the concept of an abstract or "virtual" filestore. One conceptual representation of this virtual filestore model is included in the OSI environment. In any actual operational environment, there are multiple real filestore implementations. Thus there must be a mapping between each real filestore and a conceptual virtual filestore. The nature of this mapping is considered a local issue.

Concerning the three FTAM addenda: (1) Overlapped Access deals with reading from, and writing to, different portions of a file simultaneously; (2) Filestore Management involves an extensive set of directory commands, including search, list, and change directory; and (3) Protocol Conformance involves the specification of detailed requirements to guide the FTAM implementor.

GENERAL USE:

The FTAM protocol and service specify how users on different networks can communicate about files, and transfer files, without requiring that one user know the detailed file characteristics of the other user. A generic file organization is defined for communication; elements of this virtual file model are mapped to corresponding elements of the local file system. A comprehensive set of file attributes and file activity attributes are defined, which support a large number of possible actions that can be performed on a wide variety of file types.

Some examples of applications that may use FTAM are:

- distributed data (file) management applications
- document retrieval and updating (library information services)
- specialized "messaging" systems composed of long text messages
- applications that transfer large amounts of structured data reliably end-to-end between heterogeneous systems
- large accounting and payroll applications
- large inventory control applications
- worldwide automated financial integration systems.

USE IN ACCESSIONING ELECTRONIC RECORDS:

If a variety of file storage formats are used within NARA, FTAM could prove useful as a means of managing, retrieving, and updating NARA files. The "virtual" filestore definition feature of FTAM would permit communication between the heterogeneous file systems that could exist within NARA. NARA could also employ FTAM as to transfer files on a network from outside agencies. However, care must be taken to ensure that all appropriate descriptive information is also transfered to NARA.

REFERENCES:

ISO 8571-1 (ISO TC97/SC21 N2331) - International Standard Information Processing Systems - Open Systems Interconnection - File Transfer, Access, and Management Part 1: General Introduction, Proof E, September 1987.

ISO 8571-2 (ISO TC97/SC21 N2331) - International Standard Information Processing Systems - Open Systems Interconnection - File Transfer, Access, and Management Part

1: Virtual Filestore Definition, Proof E, September 1987.

ISO 8571-1 (ISO TC97/SC21 N2331) - International Standard Information Processing Systems - Open Systems Interconnection - File Transfer, Access, and Management Part 3: File Service Definition, Proof E, September 1987.

ISO 8571-1 (ISO TC97/SC21 N2331) - International Standard Information Processing Systems - Open Systems Interconnection - File Transfer, Access, and Management Part 4: File Protocol Specification, Proof E, September 1987.

NIST Special Publication 500-150, Stable Implementation Agreements for Open Systems Interconnection Protocols, Version 1, 1988. This document can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402, stock number 003-02838-0, $12.00 (phone order @ 202-783-3238).

STANDARD NAME: Information Resource Dictionary System (IRDS)

STANDARD NUMBER: ANSI X3.138-1988 IRDS
Also: FIPS 156

STATUS: Approved by the American National Standards Institute (ANSI) as an American National Standard in 1988, and approved by the National Institute of Standards and Technology (NIST) as a Federal Information Processing Standard (FIPS) in 1989.

SCOPE:

The IRDS Standard provides specifications for modularly designed data dictionary system software used to define and maintain descriptive data, or metadata. The IRDS Standard defines "a computer software system which provides the facilities for capturing, modifying, managing, and disseminating the specifications of information and information processing resources." (ANSI X3.138-1988, Foreword).

The IRDS provides both a Minimal Schema, which supports IRDS functionality, and an example Basic Functional Schema, which provides the user with a sample schema for an application. The IRDS predefined schemas are not intended to be sufficient for all users' schema requirements; instead, the IRDS provides the user with schema extensibility, which permits each user to customize the schema of each application.

An IRD application is referred to as an Information Resource Dictionary (IRD). An IRD consists of a schema and descriptive data (or metadata) both of which may be unique to the particular application.

DESCRIPTION:

The IRDS Standard discusses "Requirements for a Conformant Implementation" and defines seven IRDS modules. These IRDS modules are:

Module 1 - Core Standard
Module 2 - Basic Functional Schema
Module 3 - IRDS Security
Module 4 - Extensible Life Cycle Phase Facility
Module 5 - IRDS Procedures
Module 6 - Applications Program Interface
Module 7 - Entity Lists

Module 1, the Core Standard, defines the Information Resource Dictionary (IRD) Schema, the IRD Minimal Schema, and two user interfaces, the Command Language Interface and Panel Interface. All additional modules are dependent on the availability of Module 1. This module also supports the users' definition of life cycle phases and user views unique to each application.

Module 2 defines the IRDS Basic Functional Schema that provides an example schema for users.

Module 3, IRDS Security, "consists of the model and functionality of an access control facility that allows organizations to restrict access to IRD and IRD Schema functionality and content. This Module modifies and extends the IRD Schema, Command Language, and Panel Interface of Module 1 to support control of access to the IRD and its IRD Schema." (ANSI X3.138-1988, p. 9)

Module 4, Extensible Life Cycle Phase Facility, "provides the basis for life cycle management of the contents of the IRD. The Module extends both the IRD Schema and the IRDS functionality of Module 1 to effect life cycle management." (ANSI X3.138-1988, p. 9) The Extensible Life Cycle Phase Facility provides the user with additional functionality to control the definition and transfer of life cycle phase metadata from one phase to another.

Module 5, Procedure Facility, "provides a mechanism for defining and executing procedures composed of IRDS commands. The Module modifies and extends the IRD Schema of Module 1." (ANSI X3.138-1988, p. 10)

Module 6, Application Program Interface, "consists of an interface to an implementation of the standard IRDS. This interface is invoked by providing Command Language syntax through the "Call" feature of any standard Language." (ANSI X3.138-1988, p. 10) This Module modifies the IRD Schema of Module 1, and requires the availability of the Command Language syntax.

Module 7, Entity Lists, "provides the capability for a name to be assigned to a list of entities, and specifies commands and panels which can be used to manipulate these lists of entities." (ANSI X3.138-1988, p. 10)

Future additions to the IRDS Standard are expected to include an IRDS Services Interface to support information exchange with Database Management Systems (DBMSs) and other systems (without requiring the availability of the IRDS Command Language) and an IRDS Export/Import File Format Facility to support schema and metadata interchange among IRDSs. Other extensions to the IRDS are also under consideration.

GENERAL USE:

IRDS specifications define software designed to support a variety of applications throughout the information system life cycle, such as data administration, structured analysis and design, configuration management, and information systems engineering.

USE IN ACCESSIONING ELECTRONIC RECORDS:

The IRDS Standard would be useful to NARA to support Information Resource Management (IRM) of accessioned records. The IRDS can support the management and cross-referencing of descriptive information about record classes and individual records. The descriptive information about accessioned records could include: the record's sources, date of accessioning, major topics or key words, the record's electronic storage format (e.g., type of text or database format), the electronic storage media, the location of the electronic storage media on which this record is stored, etc. Much of the information that describes NARA's electronic holdings would be transfered to NARA from the IRDS's maintained by outside agencies. While there exists in the IRDS standard a specification of a Minimal Schema, the actual schema information provided by outside agencies is determined by the data administration needs of the originating agency.

The IRDS could be used to support both a directory and encyclopedia of the information resources which are electronically stored as accessioned records. A set of (computer) information access screens could be used to provide users with easy access to descriptive information and, with the addition of a software program, to provide easy access to accessed records.

A prototype IRDS Panel Interface is being developed at the National Institute of Standards and Technology (NIST) to provide a screen and window interface in concurrence with the IRDS Standard. The Panel interface provides a set of screens and windows designed to make the IRDS easier to use, and to make information stored in an IRDS more accessible. An IRDS Panel Interface, and/or other screens designed specifically for NARA, would give NARA personnel easy access to descriptive information, and, if desired, additional software could be developed to provide direct access to accessed record files.

Use of the IRDS would not entail any additional descriptive information overhead, since NARA already maintains (or will soon have to maintain) the type of descriptive information to be stored in an IRDS. This descriptive information could be stored once in a central IRDS for use throughout NARA and possibly other locations. Cross-referencing information would be the sole exception to additional overhead, as NARA does not currently maintain extensive cross-referencing information on accessioned records.

REFERENCES:

ANSI X3.138-1988, American National Standard for Information Systems - Information Resource Dictionary System, Computer and Business Equipment Manufacturers Association, American National Standards Institute, 1988.

NBSIR 88-3700, *A Technical Overview of the Information Resource Dictionary System (Second Edition)*, Alan Goldfine and Patricia Konig, January 1988.

NBS Special Publication, *Guide to Information Resource Dictionary System Applications: General Concepts and Strategic Systems Planning*, Margaret Henderson Law, April 1988.

STANDARD NAME: CCITT Data Communications Networks Message Handling Systems (MHS)

STANDARD NUMBER: CCITT Recommendations X.400 - X.430

STATUS:

The Data Communications Networks Message Handling Systems (MHS) was adopted as an international standard in 1984 by the Consultative Committee on International Telephone and Telegraph (CCITT). The MHS Standard is expected to be updated in the near future, with the addition of a more thorough statement of Directory Services in the CCITT Message Handling Systems (MHS) Blue Book, to be issued in late 1989.

SCOPE:

The Message Handling Systems Standard contains a series of eight Recommendations that describe the system model, application of the model, services, protocols, and service elements of the MHS.

DESCRIPTION:

For Message Handling Systems, the following Recommendations are defined in the 1984 Red Book.

X.400 System Model Service Elements - "This Recommendation defines the **message handling (MH) services** that [network] Administrations provide to enable subscribers to exchange messages on a store-and-forward basis. Two MH services are provided. The **Interpersonal Messaging (IPM) Service** supports interpersonal communication [between an originator and recipient], including communication with existing CCITT Telex and Telematic services. The **Message Transfer (MT) Service** supports general, application-independent message transfer." (CCITT Red Book Recommendations X.400 - X.430, p.4.) (Also see Addendum).

"The expression 'Administration' is used ... to indicate both a telecommunication Administration and a recognized private operating agency." (CCITT Red Book Recommendations X.400 - X.430, p.vii.)

X.401 Basic Service Elements and Optional User Facilities - "This Recommendation overviews those [Interpersonal Messaging and Message Transfer] services and categorizes the optional user facilities of each. Locally provided service elements, for which communication with other users is not required, are not covered by this Recommendation." (CCITT Red Book Recommendations X.400 - X.430, p.39.)

X.408	Encoded Information Type Conversion Rules - "This Recommendation ... defines the encoded information type conversion rules for the MHS ... and specifies the algorithms the MHS uses when converting between different types of encoded information." (CCITT Red Book Recommendations X.400 - X.430, p.48.) (Also see Addendum.)
X.409	Presentation Transfer Syntax and Notation - "This Recommendation defines the presentation transfer syntax used by application layer protocols in message handling systems and by the document interchange protocol for the telematic services. In the architecture of Open Systems Interconnection (OSI), a presentation transfer syntax is used to represent information exchanged between application entities. This Recommendation defines a transfer syntax for various kinds of information. Each piece of information is considered to have a type as well as a value. A Data Type ... is a class of information (for example, numeric or textual). A Data Value ... is an instance of such a class (for example, a particular number or fragment of text). This Recommendation defines several generally useful types (for example, Boolean, Integer ...) from which application-specific types are constructed in other Recommendations (for example, the Message Protocol Data Units defined in Recommendation X.411). [Recommendation X.409] presents and gives an example of the intended use, standard [reference] notation, and standard [encoding] representation for each type." (CCITT Red Book Recommendations X.400 - X.430, p.64.)
X.410	Remote Operations and Reliable Transfer Server - "This Recommendation defines Remote Operations, which are used to structure interactive Application Layer protocols, such as the Submission and Delivery Protocol (P3, defined in Recommendation X.411), and describes the Reliable Transfer mechanism used between peer entities supporting the message handling protocols such as [between the] Message Transfer Protocol (P1, defined in Recommendation X.411) and P3." (CCITT Red Book Recommendations X.400 - X.430, p.95.)
X.411	Message Transfer Layer - "This Recommendation ... describes the Message Transfer Layer for MHS. MHS service elements are provided by means of the Interpersonal Messaging (IPM) and Message Transfer (MT) Services. This Recommendation defines the conceptual "layer service" provided by the Message Transfer Layer (MTL), and the peer protocols of that layer." (CCITT

Red Book Recommendations X.400 - X.430, p.128.)

X.420 Interpersonal Messaging User Agent Layer - "This Recommendation ... describes the Interpersonal Messaging User Agent Layer for the MHS. [It] defines the conceptual operation of [User Agent] UA entities within the User Agent Layer (UAL) for the Interpersonal Messaging (IPM) Service, and the syntax and semantics of the peer protocol between them." (CCITT Red Book Recommendations X.400 - X.430, p.183.) (Also see Addendum).

A User Agent (UA) is "typically, a set of computer processes (for example, an editor, a file system, a word processor) that are used to create, inspect, and manage the storage of messages. There is typically one user per UA. During message preparation, the originator communicates with his UA via an input/output (I/O) device (for example, a keyboard, display, printer, facsimile machine, and/or telephone). Also by means of these devices, the UA communicates to its user [the] messages received from the [Message Transfer Service] MTS. To send and receive messages, the UA interacts with the MTS via the submission and delivery protocol." (CCITT Red Book Recommendations X.400-X.430, p.35.)

X.430 Access Protocol for Teletex Terminals - "This Recommendation ... describes the [access model,] access protocol [and control procedures] for Teletex Terminals to [use] the MHS. This Recommendation specifies the protocol to be used by Teletex (TTX) terminals when accessing the MHS for the purpose of providing the Interpersonal Messaging (IPM) Service to their users. The IPM service elements are made available to users of the Teletex Service." (CCITT Red Book Recommendations X.400 - X.430, p.220.)

GENERAL USE:

"The establishment in various countries of telematic services and computer-based store-and-forward message services in association with public data networks creates a need to produce standards to facilitate international message exchange between subscribers to such services." (CCITT Red Book Recommendations X.400 - X.430, p.3.)

USE IN ACCESSIONING ELECTRONIC RECORDS:

MHS is not immediately useful to NARA since NARA is not sending or receiving accessioned records via data network. MHS would prove useful to NARA in the future if NARA policy is changed to include data network(s), and if records for accessioning are to be delivered to NARA via a MHS. Accordingly, NARA should remain aware of developments in this area, and records managers may find it essential to participate in the future development of, or extension of, envelope/header information.

REFERENCES:

CCITT Red Book Volume VIII - Fascicle VIII.7, Data Communication Networks Message Handling Systems, Recommendations X.400-X.430, VIIIth Plenary Assembly. Geneva 1985.

ADDENDUM:

The Red Book is expected to be replaced in late 1989 by the CCITT Blue Book. This publication is expected to show a revision of three previously defined MHS Recommendations (X.400, X.408, and X.420), and the addition of six new MHS Recommendations (X.402, X.403, X.407, X.411, X.413, X.419). The anticipated titles of these revised and additional MHS Recommendations are listed below:

X.400	MHS:	System and Service Overview
X.402	MHS:	Overall Architecture
X.403	MHS:	Conformance Testing
X.407	MHS:	Abstract Service Definition Conventions
X.408	MHS:	Encoded Information Type Conversion Rules
X.411	MHS:	Message Transfer System - Abstract Service Definition and Procedures
X.413	MHS:	Message Store - Abstract Service Definition
X.419	MHS:	Protocol Specifications
X.420	MHS:	Interpersonal Messaging System

STANDARD NAME: Database Language NDL

STANDARD NUMBER: ANSI X3.133-1986 Database Language NDL
Also: ISO 8907:1987
Also: FIPS PUB 126

STATUS:

Database Language NDL was adopted in 1986 by the American National Standards Institute (ANSI) and in 1987 by the International Organization for Standardization (ISO) and by the U.S. National Bureau of Standards as a Federal Information Processing Standard (FIPS). Derived from early CODASYL specifications, the NDL standard characterizes a class of database management systems that were very popular in the 1970's and early 1980's. These systems are the basis of a large number of existing databases in the Federal Government, but they are losing popularity for new development in favor of relational systems characterized by the SQL standard. Barring an unexpected turnabout, the NDL standard will become much less used within the next decade.

Database management systems developed in the 1970's as "CODASYL compatible" are all very close to conformance to the NDL standard, but no one product conforms exactly. With a moderate effort, all CODASYL-like products could achieve NDL conformance, but most vendors currently prefer to use their development efforts to provide SQL interfaces and/or new SQL products instead.

SCOPE:

A database language standard specifies the semantics of various components of a database management system (DBMS). In particular, it defines the structures and operations of a data model implemented by the DBMS, as well as other components that support data definition, data access, security, programming language interface, and data administration. The NDL standard specifies data definition, data manipulation, and other associated facilities of a DBMS that supports the network data model.

A database language standard is appropriate for all database applications where data will be shared with other applications, where the life of the application is longer than the life of current equipment, or where the application is to be understood and maintained by programmers other than the original ones.

DESCRIPTION:

The network data model contains two basic data structures: records and sets. A record is a collection of named data items and a set is a relationship among records. As the basic unit of data manipulation, records may be stored, erased, found, modified, and connected or disconnected from other records. Sets maintain inter-record relationships and provide logical access paths so that a user may navigate from one record to the next.

The standard provides language facilities for defining records and sets and for defining application specific views of the data. In its current form, it does not provide syntax for declaring access control restrictions; instead, such restrictions are declared in a vendor specific syntax.

The NDL standard provides a Module Language for interface to other languages. Each NDL statement may be packaged as a procedure that can be called and have parameters passed to it from an external language. As each record is found, its data item values can be passed to the accessing language as parameters.

GENERAL USE:

The purpose of a database language standard is to provide portability of database definitions and database application programs among conforming implementations. Use of a database language standard is appropriate in all cases where there is to be some interchange of database information between systems. The schema definition language may be used to interchange database definitions and application specific views. A data manipulation language provides the data operations that make it possible to interchange complete application programs.

The network data model, and thereby the NDL standard, is appropriate for highly structured, static database applications requiring rapid access along predefined paths. The network data model is less appropriate for flexible or ad hoc data retrieval and for this reason is gradually being replaced by relational and object-oriented database management systems.

USE IN ACCESSIONING ELECTRONIC RECORDS:

Because of the increasing popularity of relational database management systems, the NDL standard will likely not play a very important role in the National Archives policy for accessioning electronic records. Instead, data in existing network model databases will most often be queried through relational views, so it should be possible in the near future to interchange NDL databases using the SQL standard.

If it should be necessary to transport NDL defined databases directly, the Report on Approaches to Database Translation referenced below provides a possible mechanism using the DDF standard for actual data interchange.

REFERENCES:

ANSI X3.133-1986, American National Standard for Information Systems - Database Language NDL, ANSI, 1430 Broadway, NY 10018.

FIPS PUB 126, Federal Information Processing Standards Publication for Database Language NDL, U.S. Department of Commerce/National Bureau of Standards, March 10, 1987.

NBS Special Publication 500-115, Report on Approaches to Database Translation, by Leonard Gallagher and Sandra Salazar, U.S. Department of Commerce/National Bureau of Standards, May 1984.

STANDARD NAME: Remote Database Access (RDA)

STANDARD NUMBER:		
	ISO/IEC DP 9579	Generic Remote Database Access
	ISO/IEC DP xxxx	RDA Specialization for SQL

STATUS:

Remote Database Access (RDA) is an emerging standard under development by Joint Technical Committee One (JTC 1) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). It is specified in two pieces, a Generic RDA for arbitrary database connection and an SQL Specialization for connecting databases conforming to Database Language SQL. The initial specifications for both pieces have been completed. The formal review process will begin in late 1989 with final adoption expected in 1991.

There are no known existing implementations of these impending standards, but many SQL vendors are planning to have conforming products available before final adoption.

SCOPE:

Remote Database Access provides standard protocols for establishing a remote connection between a database client and a database server. The client is acting on behalf of an application program while the server is interfacing to a process that controls data transfers to and from a database. The communications protocols are defined in terms of Open Systems Interconnection (OSI) standards for Association Control (ACSE), Remote Operations (ROSE), and Commitment, Concurrency and Recovery (CCR). The goal is to promote the interconnection of database applications among heterogeneous environments.

DESCRIPTION:

The RDA standard provides an RDA Service Interface to an RDA Communication Element that exists both at the client site and at the server site. The RDA Communication Element converts RDA service requests into underlying ROSE, ACSE, and CCR service requests as part of an open systems interconnection.

The RDA Service Interface consists of service elements for association control, for transfer of database operations and parameters from client to server, for transfer of resulting data from server to client, and for transaction management. Association control includes establishing an association between the client and server remote sites and managing connections to specific databases at the server site. Database operations may be sent as character strings conforming to the SQL language or they may be encoded into parsed form as ASN.1 representations.

Resulting data and/or errors and exceptions are described and represented using the ISO ASN.1 standard. Transaction management includes capabilities for both one-phase and two-phase commit protocols.

GENERAL USE:

RDA is appropriate for remote access to a database in any context where OSI protocol standards for ROSE, ACSE, and CCR have already been established. It is expected that RDA will become the basis for all interconnection among SQL database management products from different vendors. Interconnection among database products from the same vendor will likely continue to use vendor specific communication and interchange forms.

USE IN ACCESSIONING ELECTRONIC RECORDS:

The National Archives is responsible for identifying and accessioning electronic records of historical value. RDA, especially with the SQL specialization, could be used by the National Archives as the client to access records remotely with the creating agency acting as server. A previously agreed "record schedule" would specify exactly the data that could be accessed. It could also be used with the National Archives as the server to provide researchers and historians automatic access to archived records.

RDA is intended to serve as a mechanism to enable limited database query from remote sites. It is not intended as a mechanism to enable transfer of entire databases between remote sites, such as from an originating agency to NARA. Thus, its use by NARA would be limited to examining limited portions of other agencies electronic holdings to determine if they are candidates for accession.

REFERENCES:

ISO/IEC 2nd Draft Proposal 9579, **Remote Database Access - Generic Specification**, ISO/IEC JTC1/SC21 N3341, January 1989.

ISO/IEC Draft Proposal xxxx, **Remote Database Access - SQL Specialization**, ISO/IEC JTC1/SC21 N3342, January 1989.

ISO/IEC, **RDA Tutorial**, ISO/IEC JTC1/SC21 N3343, January 1989.

STANDARD NAME: Database Language SQL

STANDARD NUMBER: ANSI X3.135-1986 Database Language SQL
Also: ISO 9075:1987
Also: FIPS PUB 127

ANSI X3.135.1-198x SQL with Integrity Enhancement
Also: ISO 9075: 198x

ANSI X3.168-198x Embedded SQL

STATUS:

Database Language SQL was adopted in 1986 by the American National Standards Institute (ANSI) and in 1987 by the International Organization for Standardization (ISO) and by the U.S. National Bureau of Standards as a Federal Information Processing Standard (FIPS).

The Integrity Enhancement and Embedded SQL specifications have completed the formal public review process and are awaiting formal adoption by ANSI, ISO, and FIPS. This adoption is expected to occur by summer 1989. A substantial upward compatible enhancement, called SQL2, has already been specified; it began the formal standards review process in early 1989 with final adoption expected in 1991. A second enhancement, to be called SQL3, is under development with completion expected in the 1994 time frame.

The SQL standard is very popular with a large and increasing number of conforming implementations. It is, or soon will be, the basis of definition for a majority of Federal databases and database applications involving structured data.

SCOPE:

A database language standard specifies the semantics of various components of a database management system (DBMS). In particular, it defines the structures and operations of a data model implemented by the DBMS, as well as other components that support data definition, data access, security, programming language interface, and data administration. The SQL standard specifies data definition, data manipulation, and other associated facilities of a DBMS that supports the relational data model.

A database language standard is appropriate for all database applications where data will be shared with other applications, where the life of the application is longer than the life of current equipment, or where the application is to be understood and maintained by programmers other than the original ones.

DESCRIPTION:

The basic structure of the relational model is a table, consisting of rows and columns. Data definition includes declaring the name of each table to be included in a database, the names and data types of all columns of each table, constraints on the values in and among columns, and the granting of table manipulation privileges to prospective users. Tables can be accessed by inserting new rows, deleting or updating existing rows, or selecting rows that satisfy a given search condition for output. Tables can be manipulated to produce new tables by cartesian products, unions, intersections, joins on matching columns, or projections on given columns.

SQL data manipulation operations may be invoked through a cursor or through a general query specification. The language includes all arithmetic operations, predicates for comparison and string matching, universal and existential quantifiers, summary operations for max/min or count/sum, and GROUP BY and HAVING clause to partition tables by groups. Transaction management is achieved through COMMIT and ROLLBACK statements.

The standard provides language facilities for defining application specific views of the data. Each view is the specification of database operations that would produce a desired table. The viewed table is then materialized at application execution time.

The SQL standard provides a Module Language for interface to other languages. Each SQL statement may be packaged as a procedure that can be called and have parameters passed to it from an external language. A cursor mechanism provides row-at-a-time access from languages that can only handle one row of a table at one time.

Access control is provided by GRANT and REVOKE statements. Each prospective user must be explicitly granted the privilege to access a specific table or view using a specific statement.

The SQL Integrity Enhancement facility offers additional tools for referential integrity, CHECK constraint clauses, and DEFAULT clauses. Referential integrity allows specification of primary and foreign keys with the requirement that no foreign key row may be inserted or updated unless a matching primary key row exists. Check clauses allow specification of inter-column constraints to be maintained by the database system. Default clauses provide optional default values for missing data.

The Embedded SQL specification provides SQL interface to programming languages, specifically Ada, C, COBOL, FORTRAN, Pascal, and PL/I. Applications may thereby integrate program control structures with SQL data manipulation capabilities. The

Embedded SQL syntax is just a shorthand for an explicit SQL Module accessed from a standard conforming programming language.

SQL2 significantly increases the size of the SQL language to include a schema manipulation language for modifying or altering schemas, schema information tables to make schema definitions accessible to users, new facilities for dynamic creation of SQL statements, and new data types and domains. Other new SQL2 features include outer join, cascade update and delete deferential actions, set algebra on tables, transaction consistency levels, scrolled cursors, deferred constraint checking, and greatly expanded exception reporting. SQL2 also removes a number of restrictions in order to make the language more flexible and orthogonal.

SQL3 is a forward looking SQL enhancement that intends to provide generalization and specialization hierarchies, multiple inheritance, user defined data types, triggers and assertions, support for knowledge based systems, recursive query expressions, and additional data administration tools. Standardized database export/import facilities and progress toward distributed database management standardization are also expected in the SQL3 time frame.

GENERAL USE:

The purpose of a database language standard is to provide portability of database definitions and database application programs among conforming implementations. Use of a database language standard is appropriate in all cases where there is to be some interchange of database information between systems. The schema definition language may be used to interchange database definitions and application specific views. A data manipulation language provides the data operations that make it possible to interchange complete application programs.

The relational data model, and thereby the SQL standard, is appropriate for database applications requiring flexibility in the data structures and access paths of the database. It is desirable both for applications under production control and when there is a substantial need for ad hoc data manipulation by end users who are not computer professionals.

USE IN ACCESSIONING ELECTRONIC RECORDS:

The National Archives is responsible for identifying and accessioning electronic records of historical value. The SQL Schema Definition Language is particularly appropriate for describing tables of information that may be transferred from developing agencies to the National Archives for preservation. Used with the RDA remote database access standard or with a generic data interchange standard such as ASN.1, data occurrences may also be transferred in a standard manner. A "record schedule" could specify that SQL shall be used to describe data or it could even specify the exact structure of specific tables for transfer. SQL should prove quite useful in database information transfer, but it is important that the semantic information underlying the original database also be transferred by a mechanism such as the IRDS.

REFERENCES:

ANSI X3.135-1986, American National Standard for Information Systems - Database Language SQL, ANSI, 1430 Broadway, NY 10018.

FIPS PUB 127, Federal Information Processing Standards Publication for Database Language SQL, U.S. Department of Commerce/National Bureau of Standards, March 10, 1987.

C.J. Date, A Guide to the SQL Standard, Addison Wesley, 1987.

STANDARD NAME: ISO Information Processing Systems Open Systems Interconnection Series of Virtual Terminal (VT) Standards

STANDARD NUMBER: Draft Specifications, October 9, 1986

Virtual Terminal Service - Basic Class - ISO/DIS 9040
Virtual Terminal Protocol - Basic Class - ISO/DIS 9041

STATUS:

ISO approved both VT Standards, and addendums, in 1988. ISO plans to merge the two VT documents into a single document to ~~be~~ published by ISO/IEC JTC 1/SC21 in 1989.

SCOPE:

A virtual terminal service permits users with different types of terminals and personal computers to interface transparently with a computer operating system without requiring the use of individual emulation software for each user interface. A virtual terminal service translates the signals exchanged with each user interface into a conceptual mapping of a "virtual" (or conceptual) terminal to which both the computer operating system and the user interface terminals can respond.

The terminal side user interface requires intelligence to be able to convert terminal specific information to virtual terminal information. The terminal side user interface can be supported by any type of computer hardware, ranging from a personal computer to a mainframe, depending on the application. In those instances when a (dumb) terminal uses a host to support the terminal side VT interface, a host-to-host link can be used to permit the user to access other computer systems.

Two standards are defined by ISO for Virtual Terminals: the Virtual Terminal Service and the Virtual Terminal Protocol. These user interface layers are located above the Presentation Protocol layer in the OSI model.

The Virtual Terminal Service standard provides "a model defining the interaction between users of the service ... the actions and events of the service ... [and] the parameter data associated with each primitive action and event." (ISO/DIS 9040, p.2)

The Virtual Terminal Protocol standard specifies "a set of procedures for the connection-oriented transfer of data and control information between protocol machines which implement the functions of Basic Class Virtual terminal service providers ... [and] the structure and mapping of protocol elements used for the transfer of data and control information." (ISO/DIS 9041, p.2)

DESCRIPTION:

These international Virtual Terminal standards are part of a "set of standards produced to facilitate the interconnection of computer systems. [They are] related to other standards in the set as defined in the Reference Model for Open Systems Interconnection (OSI)." (ISO/DIS 9040, p.1)

The purpose of the Virtual Terminal Service standard "is to define the service provided in the Application Layer by the Basic Class Virtual Terminal (VT) Service. The Basic Class Virtual Terminal Service is provided by the Basic Class Virtual Terminal Protocol making use of the services available from the Common Application Service Element (CASE) in the Application Layer and the Presentation Service. This international standard also defines the Basic Class Virtual Terminal Service characteristics which the VT-user may exploit." (ISO/DIS 9040, p.1)

The purpose of the Virtual Terminal Protocol standard "is to define the manner in which two protocol machines (... called Virtual Terminal Protocol Machines or VTPMs) in the Application Layer of the [OSI] Reference Model communicate in order to provide the Basic Class Virtual Terminal Service defined in ISO 9040 making use of the Presentation Layer Service and of the CASE association control service within the Application Layer." (ISO/DIS 9041, p. 1) As integral parts of this Standard, Annex A contains the [VT] State Tables, and Annex B contains the assigned ASN.1 names. Annex C, which is not an integral part of the standard, contains examples of ASN.1 encodings for Protocol Data Units (PDUs).

GENERAL USE:

In an environment in which a variety of terminals and personal computers are used for computer system access, the Virtual Terminal standards support the user's transparent mode of access to networked computer systems. These standards provide enhanced communication between systems "that support the Basic Class Virtual Terminal Service in the Application Layer of the Reference Model for OSI and which wish to interconnect in an open systems environment." (ISO/DIS 9041, p.2)

USE IN ACCESSIONING ELECTRONIC RECORDS:

It is unlikely that the Virtual Terminal (VT) Standard would be applicable to the actual process of accessioning of electronic records. VT could possibly be useful to NARA in terms of the forms to be automated and accessed by other agencies from remote locations. If a variety of microcomputer hardware will be used to support these automated NARA forms, the VT Standard could be used to standardize the user interface to these forms.

REFERENCES:

ISO/DIS 9040 - Draft International Standard Information Processing Systems - Open Systems Interconnection - Virtual Terminal Service - Basic Class, October 9, 1986.

ISO/DIS 9041 - Draft International Standard Information Processing Systems - Open Systems Interconnection - Virtual Terminal Protocol - Basic Class, October 9, 1986.

Recommendations for Document Transfer Standards and Their Integration into National Archives Policy

Attachment C:

Framework for the Exchange and Preservation of Electronic Records

Prepared for the

National Archives and Records Administration

By the

Systems and Software Technology Division
National Institute of Standards and Technology

Written by
Judi Moline

January 1989

Recommendations for Document Transfer Standards and Their Integration into National Archives Policy

TABLE OF CONTENTS

Abstract 1

1. Scope of the Paper 2

2. Introduction/Background: What is a document? 2

   2.1 The Traditional Paper Documents of Yesterday and Today 2
   2.2 Today's Dynamic Documents 4
   2.3 The Documents of the Future 4

3. Archives Requirements and Issues Concerning Electronic Documents 5

   3.1 Archives Mission 5
   3.2 Issues Concerning Electronic Documents 7
   3.3 Goal of This Project 10

4. Background - Related Standards 10

   4.1 Usefulness of Standards 10
   4.2 Document Interchange Standards 11
   4.3 Graphics and Image Standards 14
   4.4 Reference List of the Standards Discussed . . . 18

5. Model Illustrating Recommended Document Standards . . 20

   5.1 Standards Related to and Appropriate for the Archives' Document Life Cycle 20
   5.2 Links Among Document Standards 20
   5.3 Links Between the Document and Graphics Standards 20
   5.4 Proposed Use of the Document Standards 22
      5.4.1 ODA 24
      5.4.2 SGML 24
      5.4.3 DSSSL 25
      5.4.4 SPDL 25

6. Conclusion 26

   6.1 Migration to the Suggested Standards 26
   6.2 Involvement in the Standards Process 27

LIST OF FIGURES AND TABLES

Table 1: A Sampling of the Types of Written Documents . . 3
Figure 1: National Archives' Records Life Cycle 6
Table 2: Reference List of the Standards Discussed . . . 19
Table 3: Document Life Cycle and Related Standards . . . 21
Figure 2: Proposed Use of the Document Standards 23

APPENDICES

Appendix A - Types of Documents and Their Parts 28

Recommendations for Document Transfer Standards and Their Integration into National Archives Policy

ABSTRACT

The Federal Government creates a large number of documents of historical significance, ranging from text documents containing important narratives, to highly structured manuals containing important drawings, diagrams, and figures. The critical components of some of these documents may exist only in electronic form.

The following document evaluates standards for efficient text, graphic images, and technical document transfer from the creating agencies to the National Archives and recommends methods for incorporating these standards into the Archives Policy. Specifically, it recommends standards for text and technical documents that provide the capability for compound document transfer, e.g., documents containing character text, graphics, and images. It analyzes existing national and international document interchange standards, e.g., SPDL, SGML, ODA/ODIF, CGM, etc., to determine their appropriateness for this task. Procedures are recommended that could be taken by the National Archives and by agencies that create documents of historical value to achieve the most effective transfer to the Archives in a form that will facilitate future access by researchers.

1. Scope of the Paper

This paper is concerned primarily with the standards relevant to the transfer of compound documents as they currently exist or are envisioned to exist within the next few years. A compound document may contain text, graphics, and images. Secondarily, this paper suggests methods for incorporating these standards into the creating agencies' procedures and into the National Archives Accessioning Policy.

The initial parts of this paper provide an introduction to documents and to concerns which must be addressed by NARA concerning electronic documents. Then the paper presents a wide variety of document and graphic standards. After these are discussed and their functionalities delineated, the paper presents a model illustrating the use of standards in a government office. The conclusion identifies the role of NARA and the creating agencies in the transition from the world of paper documents to that of electronic documents.

2. Introduction/Background: What is a document?

2.1 The Traditional Paper Documents of Yesterday and Today

In common usage, a document is a piece of writing conveying information. It often concerns the "basis, proof, or support of something."[1] Physically, written documents are a unified whole. Binding, staples, clips, or a folder tie together the parts into what we call a document. Based on physical attributes, documents can be categorized into broad types, such as reports, papers, business documents, etc. These types can then be broken down into specific types. (See Table 1.) For each broad type of document certain generalizations can be made. For example, all reports have titles. Further, the reports are composed of paragraphs of text and they may have illustrations, charts, and graphs.

The paper documents of yesterday were typed or written out laboriously, often several times, allowing for the capture of the development of the document in the form of drafts. Each version was a document and was static. Today, with even the simplest word processor, the concept of versions can be lost. Unless versions of a document are consciously kept, the latest version usually destroys the earlier one.

Paper documents as we know them today are complex with their rich variety of contents: text with assorted modes of emphases, photographs, diagrams, tables, maps, etc. (See Appendix A.)

---

[1] Webster's Ninth New Collegiate Dictionary, Springfield, MA: Merriam-Webster Inc., 1984,371.

Generic Category	Examples
REPORT	congressional executive summary feasibility internal investigative laboratory progress research senate statistical technical test trip trouble
BUSINESS DOCUMENTS AND FORMS	invoice, etc. letters memos/memorandum minutes patent project description resume/cv sales proposal schedule
PAPER	conference position requirements
SPECIFICATION	military
STANDARD	industry international military national

Table 1: A Sampling of the Types of Written Documents

Table 1 presents some of the common documents of today. What is similar about all of them is that they are static; they are reproduced on paper and that gives them their status as documents. Two dates might be associated with such documents, the date of completion and the date of publication.

Today there is much talk of publishing/printing on demand. Although copies are generated as needed, the basic concept is still traditional. There is still the same date of completion although the date of publication might change, being the date that the document is "demanded."

2.2 Today's Dynamic Documents

Business software gives us a taste of the possible dynamic nature of documents. Report generators pull in data from the databases at the time the report is generated and the date of the document reflects that fact. If the business report is then printed, it becomes a static, written document. On the other hand, the researcher who uses electronic databases as he/she works is dealing with a moving target. Where conclusions are reached based on the statistics or trends of a particular time as reflected in the database, a "snapshot" of the relevant portion of the database must be captured. Thus, instead of referencing a published source, the researcher would reference a particular view of the database on a given date. This illustrates the dynamic nature of today's electronic documents.

Also, today we have multi-media materials. Audio-visual materials are designed to mesh with textual materials. Sometimes referred to as an electronic textbook or encyclopedia, such materials allow browsing capabilities of the materials whether it be the textual, audio, or visual portions.

2.3 The Documents of the Future

Electronic documents of the future will include the textual and audio-visual components available today. Further, they might include animated sequences, aural components, dynamic charts, tables, and graphs, etc. With more complex document processing, the parts will be pulled together by a driver program and in many senses a document will exist only if the compiled files are formatted and printed in the form of a document.

How individuals and offices will deal with the document of tomorrow remains to be seen. Today's document is already a compilation of parts stored on various devices. However, for most of us, the document is still a static, collection of materials which is bound together to reflect its unity. A document as a collection of electronic files including many media types will surely come. That is, the document of the future might consist of a text file, an audio file, and/or a graphics file. Further, the text might include a table which is created from a database at the time the document is retrieved. Such documents must be considered as plans are made for future storage and retrieval systems.

3. Archives Requirements and Issues Concerning Electronic Documents

3.1 Archives Mission

The National Archives and Records Administration (NARA) is the body responsible for safekeeping the recorded history of the Executive Branch of the United States government. Further, by agreement rather than by law, the mission of the Archives also includes legislative and court records. Because NARA requires the submission of documents and records, it is involved in the complete life of the records, the "information created, collected, processed, transmitted, disseminated, used, stored, and disposed of by the Federal Government."[2] These processes are referred to collectively as the "life cycle of the records." This life cycle concept "helps archivists and records managers to establish control over large quantities of organizational records and to ensure the efficient and timely passage of office files from one stage of activity or use to a planned, subsequent stage."[3]

The life cycle of a record can be divided into six main phases as shown in Figure 1.

Within each phase are various stages[4] or tasks.[5] (Table 3 in section 5.1 of this paper presents the specifics in the life cycle of a document along with standards indicated where appropriate.)

---

2. National Archives and Records Administration, "Memorandum to Agency Records Officers: Identifying, Describing, and Scheduling Electronic Records," NI 10.86, 1.

3. National Archives and Records Administration, (The) MARC Format and Life Cycle Tracking at the National Archives: A Study. A Report of the Archival Research and Evaluation Staff to the Archivist of the United States, Spring 1986, 11. (Hereafter referred to as MARC.)

4. Johann H. Schlichter and Leslie Jill Miller, "FolioPub: A Publication Management System," IEEE Computer (January 1988)21:1, 61.

5. Pehong Chen and Michael A. Harrison, "Multiple Representation Document Development," IEEE Computer (January 1988)21:1, 16 - 17 and 24.

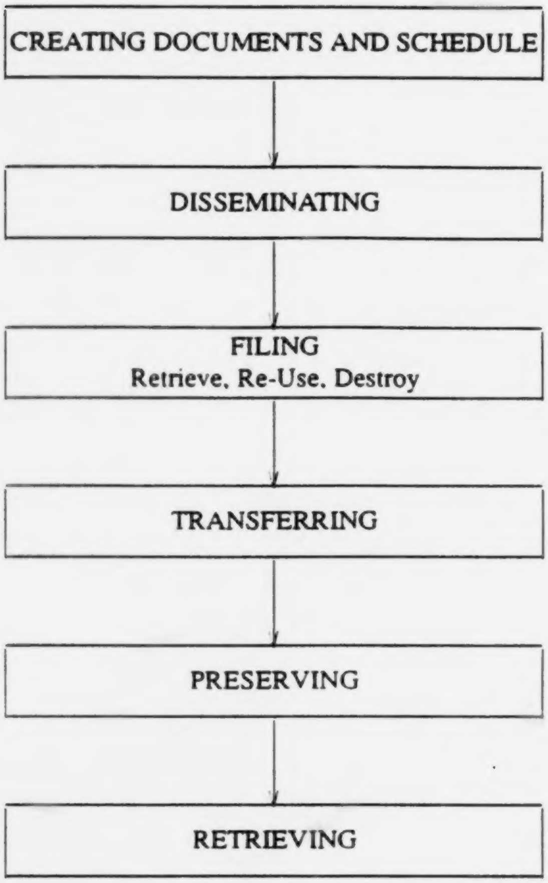

CREATING DOCUMENTS AND SCHEDULE
DISSEMINATING
FILING
Retrieve, Re-Use, Destroy
TRANSFERRING
PRESERVING
RETRIEVING

Figure 1: National Archives' Records Life Cycle

Decisions regarding the life cycle of records, regarding what records must be prepared with the idea that they might be kept, what records must be preserved, and what records must be made available, are not easily made. The Archives' staff has responsibility for the records themselves, as well as managing those records.

3.2 Issues Concerning Electronic Documents

In light of the responsibilities presented in the previous section, archivists are currently struggling with issues of technology and the need for appropriate policies. Issues related to electronic record creation, storage, and retrieval are presented in the following paragraphs.

The term document rather than record is used in much of the following discussion. Since groups of documents are dealt with as a "record," we must first have documents. Further, with future computer processing capabilities, the level of concern for NARA might drop from the record level to the document level at least for reports and other large documents. If documents were discretely identified on the record schedules, it would serve as a first step towards making it easier for searchers to find the specific archived information that they need.[6]

In the narrow sense, based on Webster's[7] definition of the verb 'archive,' all an archivist must do is file or collect materials. However, some means of accessing and retrieving materials, regardless of how cumbersome is a given, if the material is archived such that it is not destroyed.

Traditionally documents are paper based. Therefore, even if access is only given in the form of a reference to a room or to a box, retrieval is possible and as long as the symbols can be decoded, access to the material is guaranteed.

For electronic documents and records, this is no longer true. Because of changing technology, we may loose the ability to access electronic records even though the records are retrieved

---

6. See Beth Oddy's report on the Syracuse University Kellogg Project presented at the ASIS annual meeting on October 26, 1988, entitled "Making the paper archive electronic: Issues in scanning, OCR and optical disk storage." This project is striving to stretch the traditional archival methods by making effective use of computerized storage and retrieval. One of the goals is to allow the end-user access to page level retrieval.

7. Webster's Ninth New Collegiate Dictionary, Springfield, MA: Merriam-Webster Inc., 1984, 101.

intact. The ability to produce a retrieved document on a screen or in printed form is vital to its access. If there is no software capable of interpreting a file, getting the particular piece of the medium on which it is stored is not adequate.

Standardization of electronic holdings is absolutely necessary if the National Archives is to provide access to the media stored. Documents are more than a stream of letters or words. A document has form: front matter, body, and rear matter. It has layout clues: white space, paragraphs, lists, punctuation, etc. It has content clues: introduction, redundancy, etc. The content clues might be enough for a serious researcher to reconstruct the meaning but the time and effort might not be worth it.

Therefore, a whole new set of questions is raised. Are the documents to be kept in non-standard forms, thus implying that future generations wishing to read these files must first decipher them? In the case of electronic documents, this problem is exacerbated by the large variety of hardware and software available for the creation and transmission of these documents. Further, in the case of the National Archives, the archivists want to know what the historical record is and whether the key events are being captured for the historical record. These content concerns are closely related to the creation of the documents. Obviously, since the National Archives is also charged with responsibility for providing access to these materials, it must have policies which address both content and form. The content concerns the information in the records and how to access it while the form refers to the particular computer hardware and software which will allow for accessing and deciphering the information.

The importance of considering the future of electronic documents now cannot be overrated. Many people do not accept the realization that electronic documents will be the permanent record. That electronic media facilitates work is acknowledged, but that the electronic document will be the archived record is not. Currently there is a project aimed at computerizing the U.S. Patent Office. "The patent archives are stored in a cave in Pennsylvania, while the patent documents used regularly by examiners are kept in little wooden 'shoe boxes' in three buildings in Crystal City. Examiners must walk from room to room to determine if a patent already exists."[8] The question that must be asked and resolved is as follows: Must the paper documents be maintained in vaults or will the computer records suffice?

A patent is a relatively easily defined document. It consists of

---

8. Sandra Sugawara, "Computerizing Patent Office a Nightmare," The Washington Post "Washington Business" April 11, 1988.

text and image. The text and the image are tied together in that each patent consists of unique text and an image. However, if stored in a computer database, these two components might not be stored as a "document." The image might be stored in one database while the text is in another with a pointer linking them. Thus, to complicate the issue of 'What is a document?' in the future, it is likely that many documents will actually be virtual documents. That is, they will appear to be documents. In reality, they will be made up of bits and pieces from multiple sources and in various media.

There are many issues which affect the National Archives. For example, it would benefit from knowing the present extent of the use of electronic documents in government. And related, it would like to know the future extent of electronic documentation. Further, it would benefit from guidelines for the standardization of the steps in the life cycle of records so as to limit the need for long term availability of software and hardware which is necessary for machine specific electronic documents.

The MARC report of 1986 stated: "The most difficult obstacle to surmount is the lack of consistency, uniformity, or standardization in the ways in which federal agencies or, for that matter, different NARA offices collect and develop information about records."[9] Since that time "the National Archives has adopted a framework within the institution for life-cycle tracking and what is called a 'shared reference data base.' Although automated systems are still under development, the policy framework is in place."[10] However, this is still a critical problem.

As more and more individuals and organizations turn to electronic document creation with the idea of transferring them, the realization of the need for standardization will surface. It is only with careful planning that one can successfully transfer an electronic document. The software and hardware must be compatible. "The compatibility problem is so severe that publishers often choose to rekey documents that have been submitted in electronic form, and sometimes do so without notifying the authors, who are left with a false sense of security about the integrity of their texts."[11] Although NARA is unlikely to rekey documents, the possible need points up the importance of the problem.

---

9. MARC, iv.

10. Charles Dollar correspondence of October 4, 1988.

11. James H. Coombs, Allan H. Renear, and Steven J. DeRose, "Markup Systems and the Future of Scholarly Text Processing," Communications of the ACM (November 1987)30:11, 939.

3.3 Goal of This Project

The goal of this project is to identify appropriate standards for electronic document transfer.

If the materials submitted to the Archives could all conform to a single, limited mold, there would be no need for concern. No matter how poor that media, migration from one media to another could be straightforward. However, since it is not likely for such a simple solution to occur, NARA needs to establish guidelines limiting the variety of media and the systems needed to further process the records.

The Archives must choose standards and require that all the submissions conform to those standards. Since not all of the submissions will conform, there must be a policy which in the long run, at least, encourages conformance.

Thus the need is for a limited, integrated set of standards which allows sufficient flexibility while allowing the Archives to plan a minimum number of migration steps when the need arises.

4. Background - Related Standards

4.1 Usefulness of Standards

There are no recognized criteria by which to evaluate standards. Each standard has evolved through a long, arduous period of negotiation. The committees developing the National and International Standards strive to provide for the needs of their constituents. Thus, the standards are often focused towards one group, need, or preference. This may not be the most efficient of methods but it does result in broad-based work which allows for interoperability.

"Standards codify the exchange of information across an interface between two functional units and specify what is to be exchanged."[12] With the wide variety of document types, the potential variety of types of source files, and the lack of compatibility among both software and hardware, to be effective, the National Archives must impose some structure on the electronic documents which are entrusted to it. By identifying appropriate standards, the beginning of a policy or structure will be established.

---

12. David B. Arnold and Peter R. Bono, CGM and CGI: Metafile and Interface Standards for Computer Graphics, Berlin: Springer-Verlag, 1988, 3.

The following discussion of the standards focuses on the ability of each to provide interchange capabilities for compound documents.

4.2 Document Interchange Standards

Document interchange standards have emerged in response to two distinct needs. First of all, there is the need to interchange documents among workstations and tools in the office environment. Secondly, there is the need to facilitate the exchange of documents between author and publisher. The following standards attempt to satisfy those needs.

The need for a common document interchange format was first addressed by the NAVY Document Interchange Format,[13] as the representation developed at NBS (now NIST) for the U.S. Navy Department in 1984 has come to be known. The goal of this first effort was to provide manufacturers with "a common format for an agreed subset of text processing functions for interchange only. Thus, a document must be processed by a 'filter' program, developed by the manufacturers, which will do the mapping of the DIF control functions from their internal representations to DIF representations on export and the reverse on import."[14]

The usefulness of DIF is limited. It can only be used to exchange text and formatting instructions. Thus, as an interchange format, it provides transfer only of the basic document. It does not preserve the original presentation of compound documents. It only provides representation for those control functions required by the Department of Navy in 1984. At this time, DIF is proposed as a document interchange format only where no other means of interchange is possible.

The Standard Generalized Markup Language (SGML)[15] is the first International Standard which provided a basis for the interchange of documents, particularly for publishing purposes. The markup is appropriate for text only documents and provides a means for specifying other content types. However, there is no agreed upon set of content types. As a means to facilitate document interchange, SGML is limited because it does not result in a single presentation form, or even in a single set of markup tags. However, for preparing document parts for storage in databases, for marking up text to facilitate joint authoring, and for sending documents from authors to publishers, SGML works well. This is because the author describes the parts of the document as he/she creates it. The descriptive tags indicate, for example, title, paragraph, subparagraph, table, etc. There is no set of tags required by the International Standard. Rather, a document type definition (DTD) is written or adopted for each document or record type.[16] When a common DTD is agreed upon, SGML is useful for joint authoring on non-compatible systems or where the author submits his/her content to a formatter. SGML tagged document parts can be stored in a database for later re-use.

---

[13] National Bureau of Standards, Document Interchange Format (DIF), NBSIR 84-2836, 11.

[14] Ibid., 11.

[15] ISO 8879-1986 Information Processing - Text and Office Systems - Standard Generalized Markup Language (SGML) and FIPS PUB 152 Standard Generalized Markup Language (SGML).

Information Processing - SGML Support Facilities - SGML Document Interchange Format (SDIF)[17] is used to interchange SGML text files. It specifies a data structure which allows an SGML marked up document which has been developed as separate parts "to be packed into a data stream for interchange in a manner that will permit the recipient to reconstitute the separate entities."[18] For example, a cover letter might be packaged with a report for interchange. This standard is not part of the SGML standard.

The Office Document Architecture (ODA) and Interchange Format[19] has just recently become an International Standard. "ODA is a document architecture for document communication between document processing tools (editors, formatters) in multivendor office

---

[16] There are many DTDs which have been standardized for specific uses. For example, MIL-M-28001 provides two, a conforming and a non-conforming version (MIL-M-28001 Military Specification Markup Requirements and Generic Style Specification for Electronic Printed Output and Exchange of Text); the AAP has published its own (AAP Electronic Manuscript Project Standard for Electronic Manuscript Preparation and Markup Version 2.0); the DTD being developed by the Music in Information Processing Standards Committee to X3V1.8 (X3V1.8M/SD-6 of September 10, 1988 and X3V1.8M/SD-7 of August 31, 1988); and the DTD which is under development for ISO Standards (ISO/PS 2 ISO SGML Application - Specification for Standards and Technical Reports Thirteenth Draft of 1988-10-22).

[17] ISO 9069-1987 Information Processing - SGML Support Facilities - SGML Document Interchange Format (SDIF)

[18] Ibid., 1.

[19] ISO 8613-1988 Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format

systems."[20] Page layout is specified in terms of blocks in which the various content segments fit. ODA/ODIF provides for the interchange of documents in a formatted form, in a processable form which allows processing such as editing and reformatting, and in a formatted processable form which combines the two previous facilities.[21] Thus, since ODA/ODIF can be used for the interchange of documents resulting in an almost exact replication of the original document, this standard could be used for the storage of static documents. If either processable or formatted processable forms are desired, the standard would also be appropriate.

The Office Document Language (ODL)[22] is an SGML application. Since ODL is SGML-encoding for ODA documents, it allows the interchange of SGML marked-up text files. If the author submits ODL structured material with SGML tags to the formatter, the format information could be added resulting in an ODA/ODIF formatted processable form. This formatted processable form could then be sent back to the author for editing since the form allows both presentation and processing. The formatted processable form could be used for dynamic information retrieval.

An alternative to ODA/ODIF for the transfer and storage of documents is the Standard Page Description Language (SPDL).[23] The goal of the SPDL is to represent all marks on each page for fully composed, non-revisable documents.[24] Obviously, it is up to the document creators to determine whether or not a document should be revisable. However, based on the paper tradition, one might assume that once a document is "signed and sealed" it becomes a static, non- revisable entity. That the creating agency might want to retrieve that document to build on in the future is perhaps expected, but that the electronic form should be revisable within the National Archives might not be appropriate.

SPDL provides a straightforward method of representing documents which are generated by the ODA systems for printing or screen display. Likewise, it provides a capacity for similarly representing documents generated by the SGML systems whose formatting will be described by the Document Style Semantics and Specification Language, DSSSL.[25] When developed, DSSSL will provide a means for specifying the appearance of an SGML document which is machine independent. One of the stated intentions of SPDL is that document descriptions can be stored or interchanged for the purpose of presentation at other times or at other locations.

---

[20] W. Horak, "Office Document Architecture and Office Document Interchange Formats: Current Status of International Standardization," Computer (October 1985) 59.

[21] Ibid, 1.

[22] ISO 8613-5:1988 Office Document Language (ODL)

[23] ISO JTC 1/SC 18/WG 8 N561 Information Processing - Text and Office Systems - Standard Page Description Language (SPDL), 3rd Working Draft - 1988-02-19.

[24] Ibid.

The last document interchange standard to be discussed here is Electronic Data Interchange (EDI).[26] This standard was developed to allow businesses to conduct external transactions electronically with a single standard data interchange format. For example, purchase order, invoice, shipping notice, or purchase order data can be interchanged electronically. EDI is limited to business forms and thus is not appropriate for compound documents.

To conclude this section on document standards, there are two standards which provide for compound document transfer. These two standards are ODA/ODIF and SPDL. Some of the other standards discussed, as well as those which will be discussed in the following section on the graphics standards, allow for the transfer of specific types of files, but not for compound document transfer. Others are languages which allow for the description of files and/or their components. For example, SGML provides a standardized language for the descriptive mark up of text. Navy DIF allows for the transfer of text files. The two compound document transfer standards identified and discussed above will be put into a model later in this document.

4.3 Graphics and Image Standards

"Two interfaces are central to graphics standardization: the application programmer interface and the virtual device interface. Standards at both interfaces provide device independence for the user of the standard."[27]

---

[25] ISO/IEC JTC1/SC 18/WG 8 N606. This standard is in the preliminary stage of development.

[26] ANSI X12.3:1986 American National Standard for Electronic Business Data Interchange - Data Element Dictionary; ANSI X12.6:1986 American National Standard for Electronic Business Data Interchange - Application Control Structure; ANSI X12.20:1986 American National Standard for Electronic Business Data Interchange - Functional Acknowledgment; ANSI X12.22:1986 American National Standard for Electronic Business Data Interchange - Data Segment Directory. (Note: There are 11 other standards, each one for a particular type of transaction set.)

The programmer interface standards are the Computer Graphics Interfacing Techniques for Dialogues with Graphical Devices (CGI), the Programmer's Hierarchical Interactive Graphics System (PHIGS), and the Graphical Kernel System (GKS) and its three-dimensional extensions. These programmer interfaces are "typically implemented as a collection of external procedures or subroutines that a programmer can link with his application code to obtain graphical input and cause pictures to be displayed on graphical output devices."[28] All of the graphics standards for the programmer interface are initially specified in terms of their abstract graphics functionality. That is, the semantics of the functions are described but no syntax is specified.[29]

The graphics device interface standard, the Computer Graphics Metafile (CGM), "standardizes the semantics and syntax of a set of elements for the device-independent definition of pictures."[30] It is a specification for system designers and system implementors rather than application programmers. The raster component of ODA/ODIF could also be classified as a graphics device interface standard.

The Computer Graphics Interfacing Techniques for Dialogues with Graphical Devices (CGI) "is a standard functional and syntactical specification of the control and data exchange between device-independent graphics software and one or more Virtual Devices."[31] The size of CGI implies that most implementations will only be subsets of the complete virtual device. Implementations can be tailored to target environments, and yet those targeted subsets are all part of an extensible family of device interface configurations built on a common model.[32] The Programmer's Hierarchical Interactive Graphics System, PHIGS,[33] is a proposed standard which supports modeling, not just viewing, graphics objects.[34] As stated earlier, it is an application programmer's interface to a computer graphics system that controls graphics devices. It "is being designed to meet the needs of sectors of the graphics community that require a combination of hierarchical graphics data structuring a dynamic, highly interactive environment, and either 2D or 3D graphics data."[35] In short, PHIGS "is designed to be the interactive toolset for graphical model building and manipulation."[36]

---

[27] Peter R. Bono. "A Survey of Graphics Standards and Their Role in Information Interchange." Computer (October 1985) 63.

[28] Bono, Computer 64.

[29] Peter R. Bono, "Guest Editor's Introduction: Graphics Standards" IEEE Computer Graphics and Applications (August 1986) 6:8, 13.

[30] Lofton Henderson, Margaret Journey, and Chris Osland, "The Computer Graphics Metafile," IEEE Computer Graphics and Applications (August 1986) 6:8 26.

[31] ISO DP 9636:1988 Information Processing Systems - Computer Graphics - Interfacing Techniques for Dialogues with Graphical Devices

The Graphical Kernel System (GKS)[37] was the first international standard in computer graphics. GKS allows easy portability of graphics systems between different installations. "GKS divides the output of graphics into two distinct parts. The first produces output on a virtual device space called normalized device coordinates. The second allows individual workstations to interpret this virtual space in a way specified by the application program."[38] In order to provide for applications requiring 3D capability, the GKS-3D standard is being developed.[39] Although the development of GKS-3D was slow, it has recently been advanced to an ISO standard. PHIGS is the standard of choice, but for upward compatibility with GKS, GKS-3D is necessary.[40]

---

[32] Thomas Powers, Andrea Frankel, and David Arnold, "The Computer Graphics Virtual Device Interface," IEEE Computer Graphics and Applications (August 1986) 6:8, 41.

[33] ISO DIS 9592:1988 Programmer's Hierarchical Interactive Graphics System and ANSI X3.144:1988 American National Standard for Information Systems - Computer Graphics - Programmer's Hierarchical Interactive Graphics System (PHIGS) Functional Description.

[34] Peter R. Bono, "Guest Editor's Introduction: Graphics Standards" IEEE Computer Graphics and Applications (August 1986) 6:8, 13.

[35] David Shuey, David Bailey, and Thomas P. Morrissey, "PHIGS: A Standard, Dynamic, Interactive Graphics Interface," IEEE Computer Graphics and Applications (August 1986) 6:8, 50.

[36] Ibid.

[37] ISO 7942:1985 Information Processing, Graphical Kernel System (GKS), Functional Description; ANSI X3.124-1985 American National Standard for Information Systems - Computer Graphics - Graphical Kernel System (GKS) Functional Description; and FIPS PUB 120 Graphical Kernel System (GKS).

[38] Peter R. Bono, Jose L. Encamacao, F. Robert A. Hopgood, and Paul J.W. ten Hagen, "GKS - Ther First Graphics Standard," IEEE Computer Graphics and Applications (July 1982) 2:5 10.

The Computer Graphics Metafile (CGM)[41] is the first standard for a general-purpose graphical metafile. It provides a versatile and standard definition of a file for the capture, transfer, and archiving of pictorial information.[42] It does not deal with the physical record formats of the encoded data. For that, CGM depends on the groups standardizing file structure, transfer, and management. "Generality is a key attribute of CGM. It is designed for use with a wide variety of devices, applications, and systems. The same metafile can be interpreted on a low-resolution monochrome terminal, a high-resolution multipen plotter, or a raster device with high functionality."[43]

A graphical metafile (i.e., CGM) is a graphical database. It consists of a component "for generating the database concurrently with the execution of an application (the metafile generator)" and a component "for reading, interpreting, and rendering the graphical information in a metafile (the metafile interpreter)."[44] CGM is generally used to capture snapshots of GKS or PHIGS constructions and was adopted "as the picture-defining protocol of the current ISO specifications for ODA/ODIF."[45] Further, CGM pictures can be derived directly from IGES formatted databases for archiving and for inclusion in documents.[46]

---

[39] Richard F. Puk and John I. McConnell, "GKS-3D: A Three-Dimensional Extension to the Graphical Kernel System," IEEE Computer Graphics and Applications (August 1986) 6:8, 13.

[40] Mark W. Skall, telephone conversation: NIST is not planning to FIPS GKS-3D.

[41] ISO 8632-1986 Information Processing Systems - Computer Graphics Metafile for the Storage and Transfer of Picture Description Information; ANSI X3.122-1986 American National Standard for Information Systems - Computer Graphics - Metafile for the Storage and Transfer of Picture Description Information; and FIPS PUB 128 Computer Graphics Metafile (CGM).

[42] Lofton Henderson, Margaret Journey, and Chris Osland, "The Computer Graphics Metafile," IEEE Computer Graphics and Applications (August 1986) 6:8 24.

[43] Ibid., 26.

[44] Ibid., 25.

[45] Ibid., 32.

The Initial Graphics Exchange Specification (IGES) "establishes information structures to be used for the digital representation and communication of product definition data"[47] such as, the essential engineering characteristics of physical objects. IGES "is a standard format for the exchange of product definition data between different CAD/CAM systems. Data exchange using IGES is realized by software programs - the preprocessors and postprocessors that handle the access, analysis, mapping, and storage of product definition data in the CAD/CAM database and the IGES file."[48] Although sometimes categorized with the graphics standards, it does not fit in this discussion of standards. The concern here is with document rather than database standards.

The Raster Graphics Architecture[49] defines the graphics content architectures which can be used in ODA complex documents. It allows encoding according to facsimile or bitmap encoding. There is a proposed extension/addendum to include Tiled Raster Interchange Format (TRIF)[50] with the Raster Graphic Architecture.

4.4 Reference List of the Standards Discussed

Table 2 was included to relate the standards just discussed with their acronyms for reference purposes. It also seemed appropriate to include the document numbers as they relate to the various standards. Attachment A, NISTIR 88-3851, includes the full citations for the documents.

---

46. David B. Arnold and Peter R. Bono, CGM and CGI: Metafile and Interface Standards for Computer Graphics, Berlin: Springer-Verlag, 1988, 135.

47. IGES V4.0 Initial Graphics Exchange Specification, Version 4.0, June 13, 1988. (This version replaces Version 3.0, NBSIR-3359.)

48. Hans Grabowski and Rainer Glatz, "IGES Model Comparison System: A Tool for Testing and Validating IGES Processors," IEEE Computer Graphics and Applications (November 1987) 47 - 57.

49. ISO 8613-7:1988 Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format Part 7: Raster Graphics Content Architecture.

50. TTG/88-14 TRIF 2.0 Tiled Raster Graphics Content Architecture, proposed to ANSI X3V1 by the Tiling Task Group, February 1988.

Abbreviation	Name of Standard	Doc. No.
CGI	Computer Graphics Interfacing Techniques for Dialogues with Graphical Devices	ISO DP 9636:1988
CGM	Computer Graphics Metafile for the Storage and Transfer of Picture Description Information	ISO 8632:1986 ANSI X3.122-1986 FIPS PUB 128
DFR	Document Filing and Retrieval	ISO/IEC JTC1/SC 18/WG 4 N1264 ISO/IEC JTC1/SC 18/WG 4 N1265
DIF	Document Interchange Format	NBSIR 84-2836
DSSSL	Document Style Semantics and Specification Language	ISO/IEC JTC1/SC 18/WG 8 N606
EDI	Electronic Data Interchange (series of Electronic Business Data Interchange (EBDI) Standards)	
	Data Element Dictionary	ANSI X12.3:1986
	Application Control Structure	ANSI X12.6:1986
	Functional Acknowledgment	ANSI X12.20:1986
	Data Segment Dictionary	ANSI X12.22:1986
FAX	CCITT Group 4 Facsimile	Recommendations T.5 and T.6
GGCA	Geometric Graphics Content Architecture	ISO 8613-8:1988
GKS	Graphical Kernel System	ISO 7942:1985 ANSI X3.124-1985 FIPS PUB 120
IGES	Initial Graphics Exchange Specification	NBSIR 86-3359 (IGES 3.0)
	Digital Representation for Communication of Product Definition Data (Based on IGES 3.0.)	ASME/ANSI Y14.26M-1987
	Initial Graphics Exchange Specification	NBSIR 88-3813 (IGES 4.0)
ODA	Office Document Architecture	ISO 8613:1988
ODIF	Office Document Interchange Format	ISO 8613-5:1988
ODL	Office Document Language	ISO 8613-5:1988
PHIGS	Programmer's Hierarchical Interactive Graphics System	ISO DIS 9592:1988 ANSI X3.144-1988
Raster	Raster Graphics Content Architecture	ISO 8613-7:1988
SDIF	SGML Document Interchange Format	ISO DIS 9069:1987
SGML	Standard Generalized Markup Language	ISO 8879:1986 FIPS PUB 152
SPDL	Standard Page Description Language	ISO/IEC JTC1/SC 18/WG 8 N561
TRIF	Tiling Raster Interchange Format	Proposed Extension to ISO 8613-7:1988

Table 2: Reference List of the Standards Discussed (See NISTIR 88-3851 for details of each standard.)

The next part of this paper uses the document standards to build a model which might be appropriate for the National Archives and Records Administration and the agencies which create official documents.

5. Model Illustrating Recommended Document Standards

5.1 Standards Related to and Appropriate for the Archives' Document Life Cycle

Table 3 relates the document standards previously discussed (Section 4.2) with the activities covered by the life cycle (Section 3.1). The following section (Section 5.2) discusses the links among these document standards.

5.2 Links Among Document Standards

Documents marked up using SGML tags can have their format described using DSSSL and then be translated to SPDL for storage and retrieval. An ODA/ODIF document can also be translated to SPDL for storage and retrieval.

An ODA structured document can be represented for interchange by the Office Document Interchange Format (ODIF). Likewise, an ODA structured SGML document can be represented for interchange by ODL. For the same document, the ODIF and the ODL "representations are technically equivalent; a document can be transformed from one to the other without loss of information about the document constituents and attributes."[51]

Therefore, once these document standards are completed[52] the theoretical exchanges shown in the model, Proposed Use of the Document Standards (Section 5.4), will be possible. However, software applications that would perform the actual exchanges will have to be developed.

5.3 Links Between the Document and Graphics Standards

An SGML document type definition (DTD) can provide for inclusion of computer graphics in an SGML file. For example, to include IGES, CGM, or CCITT Group 4 FAX data, it is necessary to include appropriate references in the DTD. First, each graphics type must have a Notation Declaration regarding its definition. Then for each unit of data content, an entity declaration must be

---

[51] ISO 8613-5 "4 Document representations."

[52] SGML and ODA/ODIF are international standards and SPDL and DSSSL are currently being progressed in the standards' arena.

Life Cycle Activity	Related Standard(s)
CREATING THE DOCUMENT	
Authoring	SGML
Specifying Style	DSSSL
Editing/Revising	ODA/ODIF
Formatting/Processing	ODA/ODIF
DISSEMINATING	SPDL
FILING (in creating agency)	
Retrieving (to print on demand)	SPDL
Re-Using (to use parts in new document)	SGML
Destroying (discarding the file)	
TRANSFERRING THE FILE	SPDL or ODA/ODIF
PRESERVING THE FILE	SPDL or ODA/ODIF
RETRIEVING THE FILE	
Accessing	
Imaging (fixed, static)	
Viewing	SPDL or ODA (formatted form)
Printing (on demand)	SPDL or ODA (formatted form)
Dynamic Reading (context-sensitive browsing)	
Revising, Updating	ODA (formatted processable form)
Re-Using Portions (in new document)	SGML or ODA (formatted processable form)

Table 3: Document Life Cycle and Related Standards

created and included in the DTD subset.[53]

CGM (ISO 8632), referred to as GGCA in ISO 8613-8, was adopted by ODA/ODIF for geometric images (single pictorial images) and can be used to capture snapshots of GKS or PHIGS constructions. Also, with the use of Facsimile (CCITT Recommendations T.4,[54] T.5, and T.6), raster graphics are provided for in ODA/ODIF (ISO 8613-7).

Thus, the inclusion of graphics in documents is possible.

5.4 Proposed Use of the Document Standards

In order to illustrate how the standards might be used in a government office, a model has been created (see Figure 2). This model proposes that documents are created in two phases or at least with two distinct concerns. The first phase or concern might be that of the content while the second would be that of the format. Realistically, these two concerns might be handled by different people: the content might be determined by the manager while the layout is prescribed by the organization and implemented by the secretary. Thus, particularly with reports, the content specialist would create a file of the document parts with those parts clearly indicated, perhaps using SGML descriptive tags (which have been established in a DTD by the agency). The layout specialist, or a mechanical formatter, would then relate the tags to an appropriate format which may or may not include ODA/ODIF, depending on the implementation chosen.

Once the document is in final form for the particular purpose intended, it would be translated into SPDL for printing and storage. This provides a non-revisable document which can be transmitted and viewed or printed by all conforming output devices. Then, rather than destroying the SGML document, it could be placed in a database of the creating agency for later retrieval and re-use. If desired, access to this database could also be provided to others who might have need for this material.

---

[53] ISO 8879-1986 "E.2 Computer Graphics Metafile," and MIL-M-28001 Military Specification - Markup Requirements and Generic Style Specification for Electronic Printed Output and Exchange of Text.

[54] Although Group 3 Facsimile (T.4) is used in the present generation of office FAX machines, it will be replaced by the more efficient Group 4 Facsimile. Group 4 FAX is dependent on a controlled environment such as that provided by ISDN. FTS-2000 Networks A and B will both support Group 4 FAX and are scheduled to be provided between 1992 and 1993 (conversation with George Clark). Therefore, because Group 3 Facsimile is being phased out, it was ignored in this document.

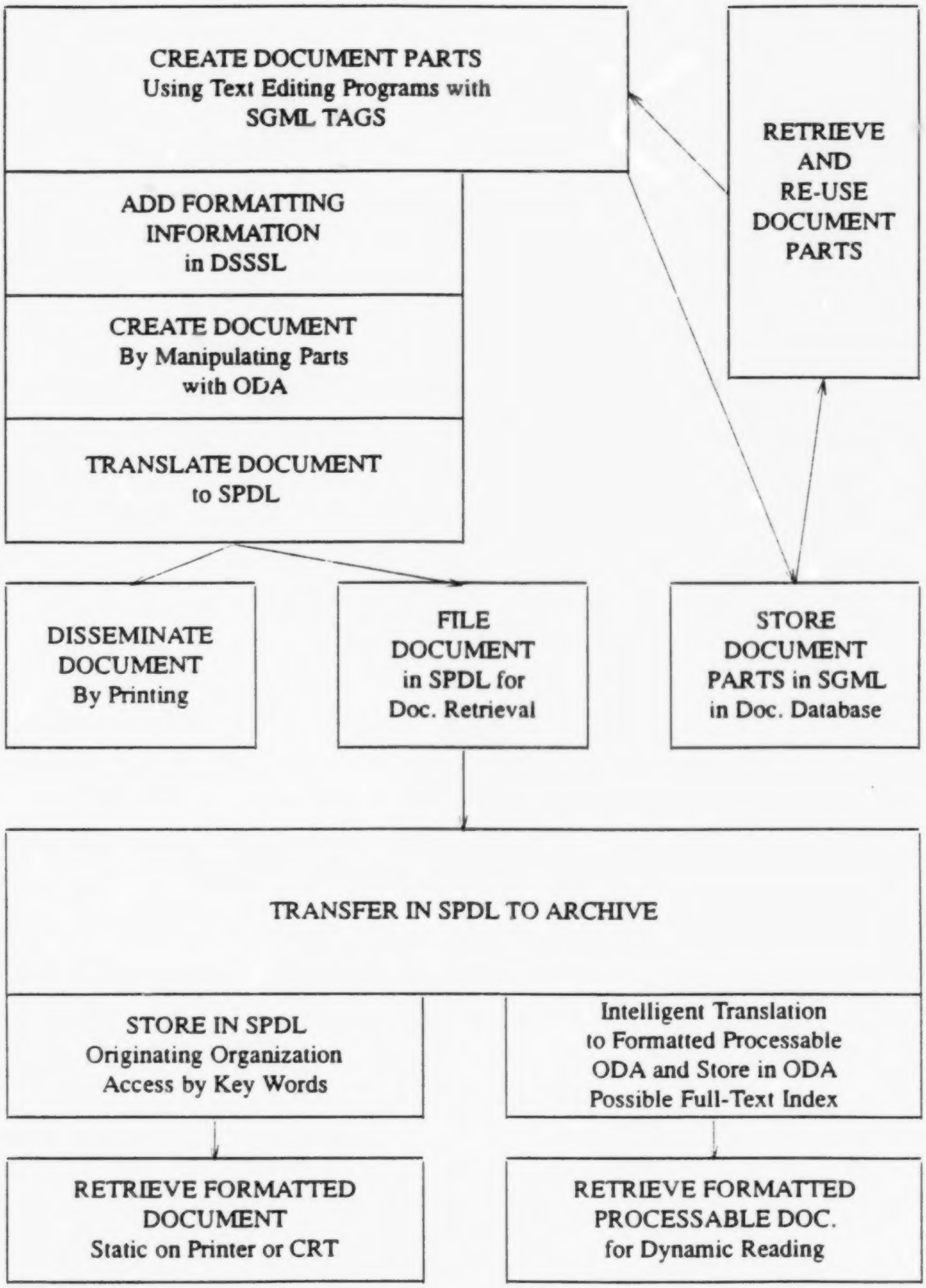

CREATE DOCUMENT PARTS
Using Text Editing Programs with
SGML TAGS
RETRIEVE
AND
RE-USE
DOCUMENT
PARTS
ADD FORMATTING
INFORMATION
in DSSSL
CREATE DOCUMENT
By Manipulating Parts
with ODA
TRANSLATE DOCUMENT
to SPDL
DISSEMINATE
DOCUMENT
By Printing
FILE
DOCUMENT
in SPDL for
Doc. Retrieval
STORE
DOCUMENT
PARTS in SGML
in Doc. Database
TRANSFER IN SPDL TO ARCHIVE
STORE IN SPDL
Originating Organization
Access by Key Words
Intelligent Translation
to Formatted Processable
ODA and Store in ODA
Possible Full-Text Index
RETRIEVE FORMATTED
DOCUMENT
Static on Printer or CRT
RETRIEVE FORMATTED
PROCESSABLE DOC.
for Dynamic Reading

Figure 2: Proposed Use of the Document Standards

It is proposed that the SPDL version of the document be considered the final document and that the documents be grouped as per the NARA policy into records that are transferred to NARA for permanent storage. Access to these documents would be based on the schedule and key words provided to NARA by the creating agencies.

Further, it is proposed in the model that when an intelligent translator from SPDL to ODA becomes available, that the SPDL be translated into ODA with full-text indexes. This would allow retrieval at the document level but might not provide for the accurate re-creation of the original document. What it would provide would be formatted processable documents which could be imported into utilities which allow dynamic reading and other hypertext-like facilities.

5.4.1 ODA

The purpose of ODA is to facilitate the interchange of documents so that text, image, graphics, and sound can all coexist, and the editing, formatting, and presentation intentions of the author can be communicated.

The Office Document Architecture (ODA) is based on the concept of structures, logical and layout. Either or both may be applied to a document. The logical objects are, for example, chapters, sections, paragraphs, and figures. The layout objects include pages, blocks, and frames (composite objects containing multiple frames or blocks). In theory, the logical and layout structures are independent of each other. The author embeds the logical structure into the document while creating and editing it. The formatter then adds the layout structure. When document objects are part of a repetitive group, they may use a generic logical structure and/or a generic layout structure. These generic structures are readily interchanged when accompanied by a document profile. A document profile is the set of attributes associated with the whole document. It includes information necessary for storage and retrieval (author, title, date) and a summary of the document architecture features used (form, content architectures, character specifications).

The advantage of ODA is the all-encompassing nature of the standard. Future extensions are planned to handle new forms of media.

5.4.2 SGML

In SGML the individual elements of a document are marked up (tagged) in such a way as to indicate their role in the document, but not the way in which they will be presented. "Because the markup codes are generic and not specific to any device, it means that a document prepared now, marked up in accordance with SGML, may be printed or otherwise published in twenty years' time, long after the present generation of hardware has been replaced."[55] Because the formatting information is not embedded in the document, the same document can be presented in many different ways by simply by altering the style sheet.

An SGML marked up document is roughly equivalent to the logically structured document of ODA. However, because the formatting is not coded concurrently with data entry, SGML can go beyond the capabilities of the machine on which the document is entered.

SGML is an easy standard for the document or database creator to use. The writer knows what the parts of his/her document are and labels them accordingly. A DTD may be created by the author, but for interchanging marked up documents, the DTD must be agreed upon. Tags must be used consistently or they will have meaning only to the document creator. There is no undue complexity and it is an efficient method for data/text entry with mark-up. However, the SGML encoding only results in a file containing the content with the parts appropriately labeled. Without an output spec relating the document parts to specific layout commands, there is no presentation document.

5.4.3 DSSSL

An SGML document needs to have format information added. DSSSL is the proposed Document Style Semantic and Specification Language for facilitating this step. DSSSL is expected to provide the means of specifying the appearance of an SGML document as it is expected to look. It is used to indicate the relationships between the logical elements expressed in the SGML DTD and the intended format, and it defines the formatting and style semantics and a corresponding machine-processable syntax.

5.4.4 SPDL

The SPDL is used to describe two dimensional images unambiguously. It provides for the representation of complex documents in their final form so that they can be output to any CRT or printing device.

A Standard Page Description Language is necessary for SGML/DSSSL documents. Further, since SPDL provides for device independent printing or viewing of stored files, translating all files (i.e., those of ODA/ODIF as well as those of SGML) into SPDL for storage would facilitate future access. SPDL preserves not only the integrity of the text, but also the layout and form of the document. It is sometimes thought that an SPDL file is significantly larger than the SGML or the ODA versions. This is not true. The SPDL file could be smaller or larger than the original[56] but the size difference probably wouldn't be significant.

---

[55] Joan M. Smith, The Standard Generalized Markup Language (SGML): Guidelines for Authors British National Bibliography: Research Fund Report 27, 1987, 1.

Storing electronic records in a standard form such as SPDL is one way to ensure future accessibility. Another way to preserve the ability to re-create the document as the author intended it, is to store not only the file, but also the software and hardware. Storing electronic records with the hardware and software on which they were produced is a cumbersome, costly way of preserving the integrity of the material. Using the Standard Page Description Language representation of the final form of a document would be a more cost effective approach.

It is proposed that the SPDL version of a document be stored in Archives. This would provide a non-revisable, full representation of the document which could be used to provide content for search purposes or to provide the content and form for retrieval purposes.

6. Conclusion

6.1 Migration to the Suggested Standards

The National Archives must build into their policy for electronic document transfer, a procedure which allows Agencies to migrate from their current procedures to one in keeping with the Archives policy.

In order to do this, NARA needs to provide guidance to the creating agencies in the form of a strategy to achieve and maintain a state of easy document interchange. Initial concern will be to formulate a strategy which facilitates the work of the creating agencies. This is not as difficult as it might appear. In many cases the agencies are floundering in a sea of choices. The model presented in this paper might be used.

Doubtless the agencies have little time or money to spend selecting the most appropriate ways to go. Guidelines to simplify their choice and to meet the requirements imposed by Archives will be welcomed. That vendors could compete by providing packages which facilitate the requirements will serve all parties concerned. A clear policy with sources of help will encourage the creating agencies to cooperate. That their help can often come from the vendors rather than from the Archives' staff, will further speed up the migration strategy.

---

[56] There are two possibilities here. The first concerns the nature of the SPDL file; it may be binary or ASCII. For example, a binary file might result in a twenty percent savings in size. The second concerns the efficiency with which the original file was prepared. For example, an SGML file constructed with unused tags will be larger than its SPDL version.

It would be expected that commercial products would cater to the needs of government agencies to include the basic functionality required to meet the guidelines established. Currently this is being done with SGML products which are capable of producing documents conforming to MIL-M-28001. (And, of course, this was done earlier when products were put on the market to support Navy DIF.)

6.2 Involvement in the Standards Process

In the short term, the National Archives should become involved with the development of the SPDL Standard. The need for a Standard Page Description Language is widely recognized and is high on the list of NIST's priorities. The group working on this Standard has invested a great deal of time and effort and would benefit from the input of technical papers providing solutions to the issues which have been raised.

In the long term, the National Archives must plan to assess the trends in technology every few years so that upgrading is done regularly. It would serve no purpose to lock into a technology with appropriate software and hardware for the long term. Rather, standards bodies must be kept alive so that today's standards evolve into those of tomorrow. Building in upward compatibility is easy if it is done on a regular basis. On the other hand, once a standard or technology is dead, it is usually difficult to upgrade it without major expense.

Appendix A - Types of Documents and Their Parts[57]

TYPES	GENERAL DOCUMENT TYPES				
	book	report	article	paper	letter
Front Matter					
cover	X	X		X	
certificate of limited edition	X				
half-title page	X				
verso of half-title page	X				
list of series editors	X				
list of books by same author	X				
list of books in same series	X				
card page	X				
imprimatur	X				
frontispiece	X				
title page	X	X	X	X	
copyright page	X				
letter or memo of transmittal	X	X			
key words			X		
abstract	X	X	X	X	
addresses for correspondence			X		X
dedication	X			X	
epigraph	X			X	
table of contents	X	X	X	X	
list of illustrations	X	X		X	
list of tables and figures	X	X		X	
errata	X				
foreword	X	X		X	
preface	X	X		X	
acknowledgements	X	X		X	
introduction	X				
opening					X
list of abbreviations	X	X		X	
glossary				X	
editorial methods	X			X	
list of contributors	X				
chronology	X				
summary	X	X			

Table of Front Matter Parts

---

[57] The data for this table is taken from the document taxonomy project being carried out at NIST by Judi Moline and Sandra Foltz. The work was instigated by Lawrence A. Welsch.

TYPES	GENERAL DOCUMENT TYPES				
	book	report	article	paper	letter
Body					
introduction		X	X	X	
chapters	X	X		X	
parts	X			X	
sections	X			X	
methods, assumptions, procedures		X	X		
theory		X			
experimental procedure		X		X	
experimental results		X		X	
main text		X			X
results and discussion		X	X	X	
conclusions		X			
summary			X		
recommendations		X			
bar charts	X	X	X	X	X
diagrams	X	X	X	X	X
displays	X	X	X	X	
drawings	X	X	X	X	
equations	X	X	X	X	X
excerpts	X	X	X	X	X
figures	X	X	X	X	X
formuli	X	X	X	X	X
graphics	X	X	X	X	
graphs	X	X	X	X	X
illustrations	X	X	X	X	
line drawings	X	X	X	X	
maps	X	X	X	X	X
matrices	X	X	X	X	X
pictures	X	X	X	X	
quotations	X	X	X	X	X
tables	X	X	X	X	X
poetry	X				
letters	X				
diaries	X				
references		X			

Table of Body Parts

TYPES	GENERAL DOCUMENT TYPES				
	book	report	article	paper	letter
Rear Matter					
acknowledgements			X	X	
appendices	X	X	X	X	
notes	X				
glossary	X	X			
endnotes	X	X		X	
literature survey		X			
bibliography, references	X	X		X	
literature cited			X		
indices	X	X		X	
closing					X
distribution list		X			X
colophon	X				

Table of Rear Matter Parts

Recommendations for Database and Data Dictionary Standards and Their Integration Into National Archives Policy

Attachment : D

Framework for the Exchange and Preservation of Electronic Records

Prepared for the

National Archives and Records Administration

By the

Information Systems Engineering Division
National Institute of Standards and Technology

Written by
Wilma M. Osborne
Bruce Rosen
Leonard Gallagher

1R016014

Recommendations for Database and Data Dictionary Standards and Their Integration Into National Archives Policy

1.0 OVERVIEW OF DOCUMENT....................1

- 1.1 Introduction....................1
- 1.2 NARA Must Know the Scope of the Archival Database Transfer....................1
- 1.3 NARA Needs a Data Administration Policy Governing Database Structure....................2
- 1.4 Accessing Archival Databases....................2

2.0 ISSUES AND CONCERNS ABOUT TRANSFER OF ARCHIVAL DATABASES....................3

- 2.1 Responsibility for Creating Archival Databases........3
- 2.2 Combining the Data and Descriptive Information........3
- 2.3 Ease of Access and Use of the Archival Databases......4
- 2.4 Selecting a Database Structure....................4
- 2.5 Data Format and Structure Considerations..............4

3.0 A SURVEY OF FEDERAL AGENCY ARCHIVAL DATABASES..............5

4.0 THE NEED TO ADOPT AND IMPLEMENT A STANDARDS PROGRAM....................7

- 4.1 Information Resource Dictionary System (IRDS).........8
- 4.2 Selecting a Standard Data Formats for Data Transfer...8
- 4.3 Selecting the Appropriate File Structure for Data Transfer....................9
- 4.4 Use of Federal Information Processing Standards (FIPS)....................9

5.0 PLANNING FOR THE TRANSFER OF SCHEDULING AND DESCRIPTIVE INFORMATION....................11

- 5.1 Transfer of Scheduling and Descriptive Information From Creating Agencies..................11
- 5.2 An Example Problem: Database Transfer................11
- 5.3 Selecting and Enhancing DBMS Capability..............12
- 5.4 Database Transfer Model....................12

6.0. RECOMMENDATIONS FOR DATABASE ADMINISTRATION POLICY........14

BIBLIOGRAPHY....................17

RECOMMENDED READING MATERIAL....................18

FIGURES

Figure 1 Data Entry/Data Presentation..........................6

Figure 2 Typical Flat File Data...............................10

Figure 3 Database Transfer Model..............................13

Recommendations for Database and Data Dictionary Standards and Their Integration Into National Archives Policy

1.0 OVERVIEW OF DOCUMENT

This document presents guidance to NARA on the planning necessary for formalizing and implementing a policy to govern the transfer of archival databases from Federal Agencies. This document provides recommendations to NARA on the selection and use of database standards.

1.1 Introduction

As organizations seek to identify data management systems that satisfy their requirements, the number and variety of installed DBMSs continues to grow. Today, most Federal Agencies, hereafter referred to as "creating agencies", are performing data processing and management using a variety of operating and database systems to support their burgeoning data management and administration requirements. It is not unusual for database information to be contained on a variety of hardware and software systems in different departments, and even different physical locations. While the vendor community is moving in the direction of standardizing relational DBMSs through application of the SQL standard, the individual differences between vendor systems, plus the decision by some agencies to modify vendor software (or have their own unique systems developed) still results in great diversity within the Federal government. Thus, attempting to make transfers of data and information from one system to another can be an arduous task. While this problem can be reduced for those transfers involving systems which employ the SQL standard, the transfer of archival databases from the creating agencies to NARA will also involve moving data and information from one database system to an entirely different type of database system. Much of the difficulty in transferring the archival databases will be ensuring that the associated, descriptive information about the data, which is generally maintained separately, can be readily accessed with the data. Thus, a comprehensive data administration policy supported by recognized standards is essential.

1.2 NARA Must Know The Scope of the Archival Database Transfer

Prior to assessing the feasibility of archiving databases provided by the creating agencies, the scope of the transfer of data and information to be archived must be defined. One of the key issues to be addressed is the universe of the database information to be archived. NARA should conduct a survey of the creating agencies archival requirements to determine the scope of the data to be transferred. Because of the possible multiplicity of database structures and database management systems used by the creating agencies, NARA must have some understanding of creating agencies' database systems including how they create, generate, store, access, retrieve, and manage the databases to be archived. By having this understanding, NARA can seek to ensure that important relationship information is not lost in the transfer of database information into the format to be archived.

1.3 NARA Needs a Data Administration Policy Governing Database Structure

NARA needs to establish a data administration policy which specifies the acceptable database structure for transfer of data to NARA. The policy should specify procedures and responsibilities for both NARA and the creating agencies regarding the transfer, acceptance, storing, retrieving, managing, accessing, and distributing of archival database information.

It is essential that NARA receive all needed information about the structures of both existing and planned databases that may be accessioned. NARA's policy should specify how to handle these databases; whether the databases will be available on-line to the creating agencies, researchers, and others, or whether they will be stored on tape, hardcopy, etc.. NARA's policy should require information from the creating agencies on both current and planned database structures so as to ensure that NARA has the capacity and necessary resources to manage accessioned databases.

No matter what physical media is chosen for transferring the archival databases, the input to NARA from the creators of the databases must consist of the: 1) actual database in a flat file format; 2) Information Resources Dictionary System (IRDS) description of original database schema, along with other information about the database such as source of original data; and 3) description of the flat file (perhaps another Information Resources Dictionary (IRD) which conforms to the IRDS standard or a description of the flat file using SQL Schema Definition Language). Flat files are discussed in detail in Section 4.

1.4 Accessing Archival Databases

If NARA decides to make actual archived databases available on-line, in a structure that is equivalent to the original database, they must identify a way to represent any database structure under whatever DBMS(s) are available at NARA. This would be a formidable, if not impossible task under today's technology. It may be possible to tie the structure of all archived databases together if NARA finds a common thread during their survey of the creating agencies' database structures. However, other then the fact that all of the databases may be transferable as normalized flat files, this is unlikely. While all archived databases may be transferred as normalized flat files, this flat file structure does not lend itself to performing complex adhoc queries, and thus would not be suitable to serve as the common thread needed.

Even if archived databases are available on-line, the descriptive information about these databases would still be maintained under the same recommended data administration function as used for all other NARA holdings. (See the report entitled "Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records".)

2.0 ISSUES AND CONCERNS ABOUT TRANSFER OF ARCHIVAL DATABASES

There are a number of issues that must be addressed by the creating agencies prior to planning for the transfer of archival database information. It is important that NARA not only identify these issues, but discuss the priority in which they should occur, as well as the implications of addressing or not addressing these issues. How will the creating agencies provide uniform data and information that can be readily utilized by NARA? Decisions made will have to be based on what is achievable and practical versus what is desirable. NARA and the creating agencies will need to develop a strategy for deciding on responsibilities, approaches for combining data and descriptive information, structure of databases, etc.

2.1 Responsibility for Creating Archival Databases

One of the lessons learned on previous major projects is that every group and every organization that will be affected by the decisions relating to the transfer of data and the associated descriptive information, or at least someone who represents the groups' interests, must be involved in the planning for the transfer.

While the creating agencies that have responsibility for providing the archival database information are best equipped to define the types of data and information to be archived, NARA has a responsibility to provide guidance in this area. In order to cover the widest possible range of databases, it is recommended that NARA seek to utilize normalized tables, assembled in a flat file format.

2.2 Combining the Data and Descriptive Information

One of the most challenging problems of supplying descriptive information is determining relevance. That is to say, how much descriptive information about an entity is adequate for understanding both the purpose and utility of specific data items. Certainly, much of this class of information is not usually stored with the data or in a contiguous manner. Thus, documenting the goals and objectives of the database content is a necessity in order for that data to be valuable in the future.

2.3 Ease of Access and Use of the Archival Databases

Another issue that should be considered is the ultimate use of the data. Although archival databases are not likely to be used frequently, when needed, they must satisfy the requirements of every potential user. The data and the associated, descriptive information should be readily accessible, retrievable, and reportable, whether stored on-line, on disk, or on some other medium.

2.4 Selecting a Database Structure

NARA and the creating agencies need to determine a suitable structure(s) for archival databases. Government and industry seem to be moving toward the adoption of relational databases because of the flexibility that they provide. For example, a relational database allows logical data to be organized as tables without embedded pointers or structural elements. Indexes and keys are used to navigate through the tables (database). The logical structure of a relational database lends itself more readily to maintenance than do the other database management system (DBMS) approaches. The separation of logical information from the physical structure is useful for optimal data design, since data design can take place undisturbed by physical structure requirements. Thus, the addition and deletion of tables and rows can be accomplished with little affect to existing applications or data. This type of database structure is extremely useful for organizations that have data requirements which frequently change. Even many non-relational database systems are now providing relational views for standardized access. It also appears that the most recent Object-Oriented and Knowledge-based systems are making their static data available through relational views, especially SQL interfaces.

2.5 Data Format and Structure Considerations

Unlike paper records which are fixed, electronic records can be stored in different ways which will not change the semantics, or meaning, of the information, but would change the syntax, or format, of the information. Since NARA will not be receiving actual "databases" in their normal structured form (syntax), there exists the possibility that NARA might wish to actually specify format and structure changes for individual elements within given databases in order to simplify future access to the data. In making this decision NARA will have to consider both output format considerations along with ease of retrieval. Also NARA must consider that if this decision is made, then NARA's electronic holdings would be slightly different from the original databases, although the actual information content would remain the same. Quite often the best choice for output presentation will cause increases in retrieval processing. The reverse of this situation is that the best choice of format for retrieval purposes often increases the processing needed to provide acceptable output presentation. The easiest choice is to select an input form that matches the desired output format (see Figure 1.). An example might be date: month, day, year, in the order- 02/10/89. It is very likely that the date will be the object of a search operation when selecting records for report generation. It may be desirable to select records on the basis of whether the date precedes or follows a given date, or if the date falls within range. It may be desirable to select records on the basis of specific day, month, or year of event. For maximum flexibility, it would, therefore, be desirable to address individually each component, day, month, and year. Does that mean that the data becomes three fields and not just one? Perhaps some consideration should be given to selecting a DBMS that has the more advanced feature of sub-fields which permits doubling up of these requirements.

Other issues to be considered include trade-offs in storage utilization versus establishment and execution ease or availability of variable-length field capability. These concerns may not be critical since in most large DBMS(s), a more comprehensive approach to these problems is possible. For example, a large DBMS may have 10 to 15 options for format of date presentation and 7 to 10 options for presenting names. Each of these options can be invoked in any order so that the user has a factorial number of variations.

While the issues discussed in this section might not appear to be important until after data has actually been transferred, by considering them prior to establishing the rules for database transfer, it may be possible to limit the amount of reformatting or restructuring that NARA would have to perform on transferred data.

3.0 A SURVEY OF FEDERAL AGENCY ARCHIVAL DATABASES

An assessment of what is currently being used: software (e.g., programs, documentation, databases, tools), and hardware (e.g., operating system, hardware configurations) is essential in order to determine the feasibility of transferring databases to be archived. There are several approaches that could be used for conducting such an assessment. These approaches include:

- surveys/questionnaire
- meetings with agency management officials
- workshops for both the technical and management staff.

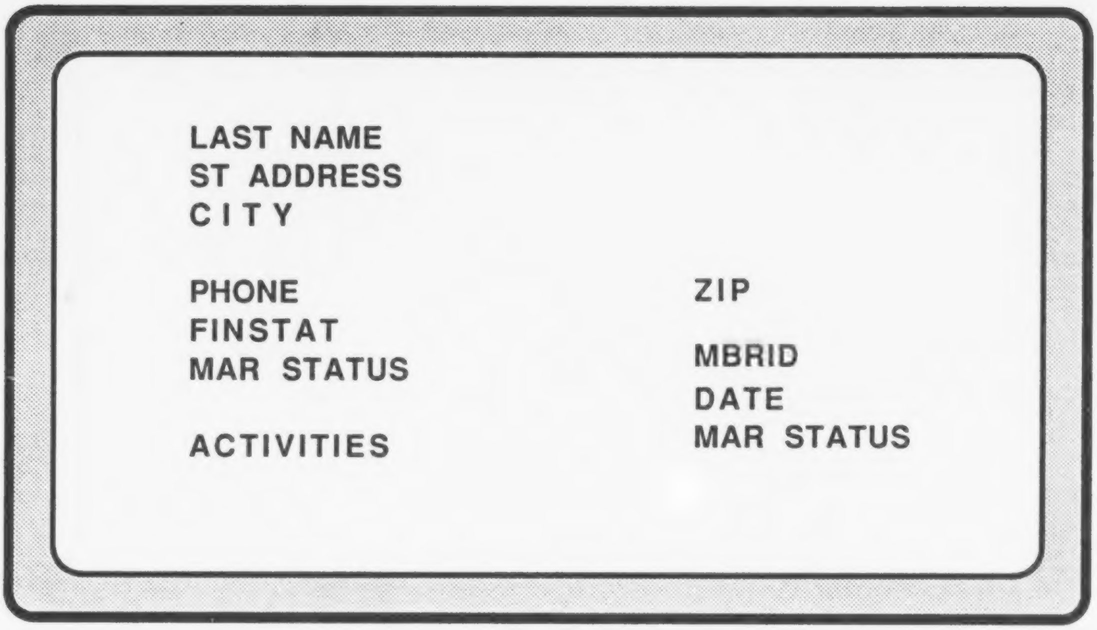

LAST NAME
ST ADDRESS
CITY
PHONE
FINSTAT
MAR STATUS
ACTIVITIES
ZIP
MBRID
DATE
MAR STATUS

Figure 1. Data Entry/Data Presentation

Of these, however, the most effective approach is the survey or questionnaire.

Since each of the creating agencies have different goals, missions, and objectives, their data requirements are also different. However, there are many commonalities in both the kinds of data and information the creating agencies generate, collect, manage, use, and store. There are also some commonalities in the way these organizations decide to perform these activities. The greater the number of commonalities, the easier it will be for the creating agencies to provide data and descriptive information in a uniform format.

There is also a possibility that there are sets of functions, attributes and characteristics common to many databases used to support mission goals, objectives, and accomplishments. While any of these sets could be divided into sub-sets, many would be recognizable to multiple organizations. A survey of the creating agencies archival database requirements should address, but not be limited to the:

- data to be archived (criteria used to select)
- amount of data to be archived
- domain/attribute sharing data to be archived
- procedures and other related data to be archived
- storage media requirements
- database systems requirements
- operating systems requirements
- various uses of data
- internal data storage format (e.g., EBCDIC, ASCII)
- data organization (e.g., centralized, decentralized)
- level of activity
- number of changes
- frequency of change to data
- percentage of data changed weekly, monthly, yearly
- life of data
- current structure of data
- future structure of data
- interfaces
- tools (conversion, restructuring, reuse)
- historical value/importance.

4.0 THE NEED TO ADOPT AND IMPLEMENT A STANDARDS PROGRAM

Standards, as used here, refer to a consistent mode of operation, use of techniques, selection of human resources, etc. When discussing standards, care must be taken to ensure that everyone is aware of level and applicability of the standard being discussed.

4.1 Information Resource Dictionary System (IRDS)

Since there is no single database structure that is right for all applications, even those within one organization, some common means of describing all archived databases must be utilized. The IRDS provides its users with the freedom to describe the structure utilized by each individual application database. The IRDS specifies database descriptive information in terms of entities, relationships, and attributes. An entity can represent a person, concept, event, data element, or quantity, but it is not the actual application data. A relationship is an association between two IRD entities (e.g., the Payroll Record "contains" Social Security Number). Attributes are the representation of properties of an entity or a relationship. The IRDS contains facilities to create, delete, modify, or copy entities and relationships. It is equipped with an output facility which can produce reports and prepare queries. A detailed discussion of the IRDS features is provided in NBSIR 88-3700, A Technical Overview of the Information Resource Dictionary System (Second Edition).

In March 1989, the IRDS was approved as FIPS 156. Now that it is a FIPS, the IRDS can be used by NARA as the base standard on which the maintenance and transfer of descriptive information, about archived databases, can be based.

4.2 Selecting a Standard Data Format for Data Transfer

In order to transfer to NARA the possible myriad of database information, it is essential to employ standards. The two most widely used standard data formats for information transfer today are the Data Descriptive File (DDF) and the Abstract Syntax Notation One (ASN.1). Both of these standards are intended to facilitate moving data structures from one computer system to another, independent of make and architecture.

The DDF specifies medium-independent and system-independent file and data record formats for the interchange of information between computer systems. The DDF is a file containing a data descriptive record and its companion data records.

The international standard, ASN.1, specifies a notation for an abstract syntax definition. The encoding rules specify the representation, during transfer, of the value of any ASN.1 type. ASN.1 notation consists of a sequence of characters from a character set specified by ASN.1.

The question that must now be resolved is: can either of these standards be employed as the single method by which all eligible Federal government databases can be transferred to NARA, or must both be utilized? The total inventory of databases within the Federal government uses technology that ranges from obsolete to state-of-the-art. Most of the technology, however, is quite old. Organizations that normally produce a substantial percentage of their output as flat sequential files might prefer to provide archived database information in DDF format. Other organizations that are concerned with communication protocols might prefer to use ASN.1 format.

The three possible choices that NARA can make between these two standards are: select DDF as the standard for all database transfer; select ASN.1 as the standard for all database transfer; allow organizations to utilize either DDF or ASN.1, depending upon the structure of the original database being archived. In order to make the needed decision, NARA should take into account both the current population of databases to be archived, and what the future of these databases, and any new databases, is expected to be. Based on the direction of technology, it would appear that, even though many current databases would be easier to transfer under DDF, in the future, ASN.1 will be the standard of choice. No matter which of the above three choices NARA selects, the use of the IRDS as the medium for documenting all archived databases is appropriate.

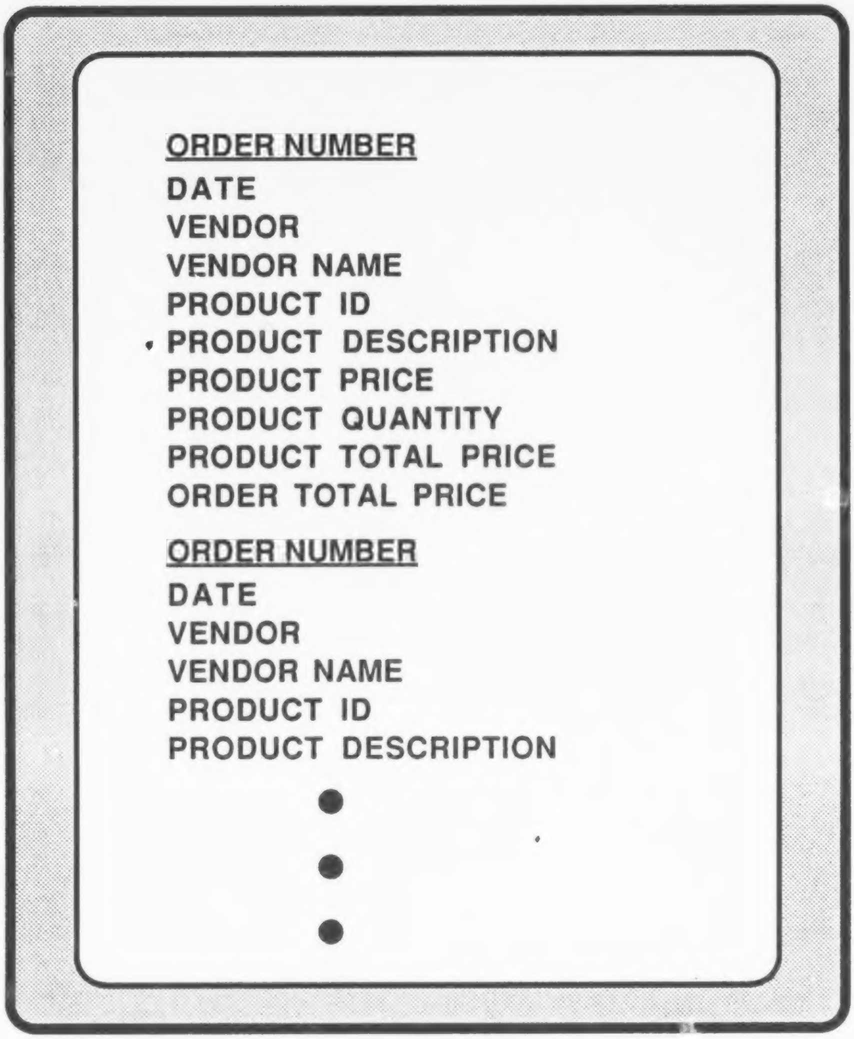
4.3 Selecting the Appropriate File Structure for Data Transfer

What must the creating agencies do in order to provide archival database information in a form that NARA can process? Even after the decision has been made concerning the standard to be utilized for database transfer, there are additional decisions that must be made concerning the file structures of databases to be archived. One of the first activities the creating agencies must accomplish is to get the data in line to be transferred to some "standard" form without changing the meaning of the data or its domain, attributes, interfaces, and associated procedures.

After careful study of the alternatives, it appears that the most effective form in which database information could be provided to NARA would be as a linear or flat data file. A flat data file may be compared with a table (See Figure 2.), or it could be defined as follows: A flat file is a sequence of records each having a fixed number of fields and containing exactly one elementary data value for each field. The alternatives, i.e., hierarchies and networks, introduce unacceptable complexities.

4.4 Use of Federal Information Processing Standards (FIPS)

One major consideration that should be part of establishing the procedures and rules for database transfer to NARA should be the application of FIPS. In the short term, NARA should seek to select and apply applicable FIPS in the establishment of their transfer policy. Then, in the long term, a joint program by NARA and NIST to establish a specific FIPS covering the archiving of database information could be developed. Once this FIPS is



ORDER NUMBER
DATE
VENDOR
VENDOR NAME
PRODUCT ID
PRODUCT DESCRIPTION
PRODUCT PRICE
PRODUCT QUANTITY
PRODUCT TOTAL PRICE
ORDER TOTAL PRICE

ORDER NUMBER
DATE
VENDOR
VENDOR NAME
PRODUCT ID
PRODUCT DESCRIPTION

●

●

●

Figure 2. Typical Flat File Data

developed, then Federal government organizations not in compliance could be required to fund the translation of their archive databases to the FIPS format.

5.0 PLANNING FOR THE TRANSFER OF SCHEDULING AND DESCRIPTIVE INFORMATION

Critical questions have to be addressed by both NARA, which has responsibility for receiving, maintaining, and managing the archival database and associated descriptive information, and by the creating agencies. A few of the more difficult ones that have to be addressed during the early planning phase are:

- Who within NARA decides which data format is appropriate?
- How will NARA enforce the requirements on the creators of data?
- Who in NARA will be responsible for ensuring that the database contains what is expected? Will the same procedures that are currently in use for paper records be sufficient?

5.1 Transfer of Scheduling and Descriptive Information from Creating Agencies

Scheduling and other descriptive information about records is produced by the creating agencies, in conjunction with NARA, to be transferred to NARA, along with the specified records. Once received by NARA, this scheduling and descriptive information is reviewed and evaluated by NARA staff, who may augment this information as necessary. After this scheduling and descriptive information is reviewed, it is stored to maintain information about the record content and structure, and information for record access and retrieval.

5.2 An Example Problem: Database Transfer

The importance of descriptive information about record content and structure is particularly obvious in the process of exchanging database records. One of the difficulties encountered by NARA is the lack of an adequate means of receiving and transferring database information without the loss of meaning. At this time, meaningful database information can only be adequately transferred: (1) in the limited format of reports and query results, through query languages such as SQL; (2) from active database to active database (i.e., where active database denotes the presence of both computer hardware and DBMS software); or, (3) when database information is represented in a flat file of data, and is accompanied by complete descriptive information of the structure and other features of the flat file and the original database.

5.3 Selecting and Enhancing DBMS Capability

If NARA elects to upgrade, enhance, or select one or more DBMSs in order to facilitate making archived databases available online, NARA should consider one which:

- has comprehensive facilities
- supports the entire spectrum of users
- provides or accommodates a wide variety of effective application development aids
- is compatible with current or planned system architecture
- effectively communicates with and supports other DBMS(s)
- directly accesses the existing data and support tools
- provides strong connectivity between systems
- has widespread industry acceptance and longevity.

5.4 Database Transfer Model

In order to help clarify the database transfer concept, a model (see Figure 3) has been created. This model, while extremely limited in scope, illustrates how a database transfer to NARA could be prepared utilizing such standards as SQL and the IRDS. The solid arrows in Figure 3 represent the actual movement of data, while the "shadow" arrows represent the act of reviewing, or utilizing, information from the indicated source in order to accomplish the indicated action.

Once a decision has been made to transfer a database to NARA the originating agency would take the following actions:

1. Produce a flat file version of the database using SQL.

2. After reviewing both the flat file and the organizational IRDS, produce a description of the flat file's syntax under the IRDS. Produce an IRDS Export version of this information.

3. Extract, from the organizational IRDS, the schema description and the semantic information about the original

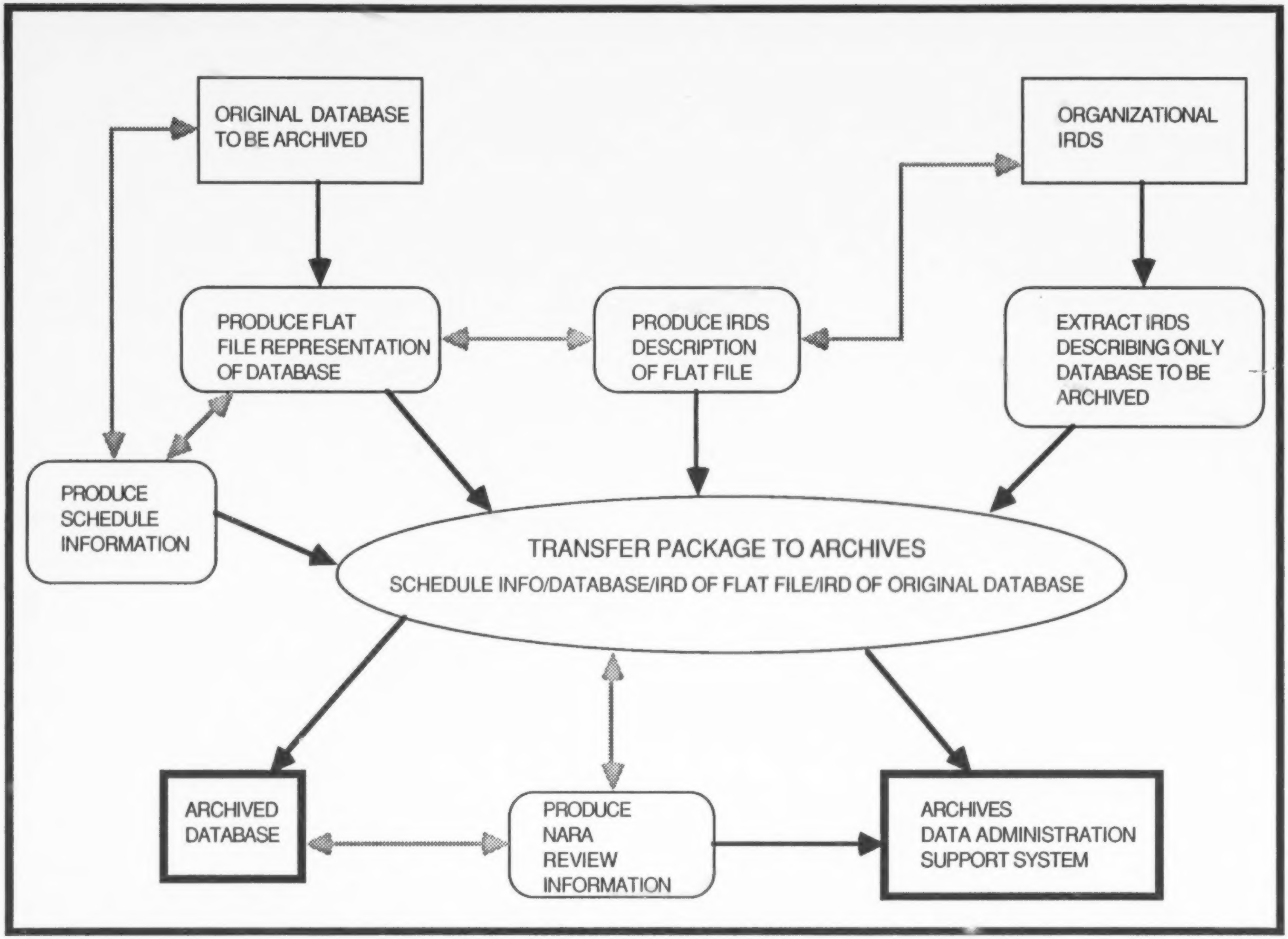

ORIGINAL DATABASE TO BE ARCHIVED
ORGANIZATIONAL IRDS
PRODUCE FLAT FILE REPRESENTATION OF DATABASE
PRODUCE IRDS DESCRIPTION OF FLAT FILE
EXTRACT IRDS DESCRIBING ONLY DATABASE TO BE ARCHIVED
PRODUCE SCHEDULE INFORMATION
TRANSFER PACKAGE TO ARCHIVES
SCHEDULE INFO/DATABASE/IRD OF FLAT FILE/IRD OF ORIGINAL DATABASE
ARCHIVED DATABASE
PRODUCE NARA REVIEW INFORMATION
ARCHIVES DATA ADMINISTRATION SUPPORT SYSTEM

FIGURE 3. DATABASE TRANSFER MODEL

database to be archived. Produce an IRDS Export version of this information.

4. After referencing the original database and the flat file representation, produce the appropriate schedule information.

5. Combine the following to produce the transfer package to be provided to NARA: the schedule information; the flat file database; the Export version of the IRDS describing the flat file; the Export version of the IRDS describing the original database.

NARA, upon receiving this transfer package would take the following actions:

1. Separate the database flat file and store it in the appropriate location.

2. After reviewing the archived database and transfer package, produce any needed NARA review information (e.g. location of archived database, etc.) and add this information to the NARA Data Administration support System.

3. Use the IRDS Import facility to transfer the description of the flat file to the Data Administration Support System.

4. Use the IRDS Import facility to transfer the description of the original database to the Data Administration Support System.

6.0 RECOMMENDATIONS FOR DATABASE ADMINISTRATION POLICY

A number of issues that relate to the transfer, access, management, and output of databases are discussed in both the Framework report and in this report. This section provides guidance to facilitate the transfer of archival database data and information from the creating agencies to NARA. It is intended to assist NARA in not only the transfer of archival databases, but in establishing and implementing a data administration program. After a review of NARA database management functions, we have concluded that:

1) NARA must establish a set of data administration policy guidelines to be provided to outside agencies. Since each creating agency has different missions and goals, and different approaches for accomplishing those missions and goals, the data administration policies of each organization are probably different. An understanding of the capabilities and constraints of the creating agencies is essential before NARA can attempt to formulate guidelines for the transfer of archival databases from the creating

agencies. The overriding considerations, however, are the ease with which the data can be transferred to NARA repositories, managed by NARA database administration, and retrieved from the NARA repositories.

2) NARA should require that the descriptive information (IRD) about the data be complete. Since file and data manipulation is dependent on structure, either great flexibility is required or total definition of the file structure must be available.

   If the database is transferred in a flat file format, the descriptive information about the schema structure of the database and its query and report structures will be missing. This, and other domain information must be captured, transferred, and maintained along with the flat data file. The capture, transfer, and maintenance of both the flat file and the descriptive information is essential to the understanding and use of a transferred database.

3) NARA should adopt and use the IRDS as a tool for ensuring the integrity of the archival databases. This standard, which is discussed in detail in NISTIR 88-3700, provides a mechanism for control over the input and output from the database system.

4) NARA must conduct a survey to determine the types and structure of the archival databases the creating agencies plan to transfer. One of the difficulties faced by NARA is that a large number of the databases used by the creating agencies differ not only in design, structure, and function, but in processing method, definition of elements, types of linkages and pointers as well. Because many of the systems and much of the technology employed to generate and maintain these databases are both antiquated and cumbersome, new or different DBMSs may be selected. Depending on the DBMS chosen, the structure or format of the data could change significantly.

5) NARA must have the capacity to manage the descriptive cross-referencing information (e.g., linking the flat file data part of the record to the descriptive part of the record), if the data and descriptive information is to be available to researchers. Cross-reference information is that information that links the content of an actual database with the appropriate descriptive information about that database. This capability is critical for the retrieval and use of this information from an archival database. The data records and the descriptive information about those data records contain different types of information, and are generally stored separately, sometimes in different physical

locations. Therefore, the methods of accessing the data and the associated descriptive information will be different.

6) NARA must decide, in conjunction with the creating agencies, what data structure and format is the most appropriate for storing, accessing, and maintaining the archival databases.

7) NARA should assign specific personnel responsibility for managing the various sections of the archival database. The integrity of the database is extremely important and must be assured at all times by proper and timely checks designed to isolate discrepancies and inconsistencies. All responsibility for manipulation of the database should reside within a single authority within NARA. Gathering data for the creation of an archival database to be transferred can be a significant task and requires the cooperation of both the suppliers and users of the data and its descriptive information. Consequently, the data collection and manipulation procedures should be addressed in NARA's data administration policy.

8) NARA should implement the following ten-step plan to ensure that the data administration policy accomplishes what is intended.

   a. Establish and define the goals and objectives.

   b. Identify all areas that have to be coordinated.

   c. Identify, to the extent possible, the present and the future environments.

   d. Identify alternatives for accomplishing objectives.

   e. Determine the benefits and, to the extent possible, the costs of each alternative.

   f. Evaluate alternatives (standards, structures, etc.,) using criteria established in conjunction with the creating agencies

   g. Conduct a "what if" analysis.

   h. Review the results.

   i. Adopt the "best suited" approach for meeting agreed upon objectives and goals.

   j. Conduct one or more pilot projects in this area, and then revisit steps 3 through 10 (c.-j.) to modify them, if necessary, prior to large-scale implementation.

BIBLIOGRAPHY

[1] Goldfine, A. and Konig, P., A Technical Overview of the Information Resource Dictionary System (Second Edition), NBSIR 88-3700, January 1988.

[2] American National Standard for Information Systems, Information Resource Dictionary System (IRDS), ANSI X3.138-1988, ANSI, Inc., New York, New York, 1988.

[3] Moline, J, Document Interchange Standards: Description and Status of Major Document and Graphics Standards, NISTIR 88-3851, NIST (formerly NBS), September 1988.

[4] Law, D. and Longworth, G. Systems Development: Strategies and Techniques, NCC Publications, Oxford Road, Manchester M1 7ED, England, 1987.

[5] American National Standard for Information Systems, Specification for a Data Descriptive File for Information Interchange, ANSI/ISO 8211-1985, FIPS PUB 123, New York, New York, February 1986.

[6] Dayal U. and Hwang, H. Y., "View Definition and Generalization for Database Integration in a Multidatabase System", IEEE Transactions on Software Engineering, Vol. SE-10, Number 6, pp. 628-644, November 1984.

[7] Attar, R., Bernstein, P. A., and Goodman, N., "Site Initialization, Recovery, and Backup in a Distributed Database System", IEEE Transactions on Software Engineering, Vol. SE-10, Number 6, pp. 645-649, November 1984.

[8] Gallagher, L. and Salazar, S., Report on Approaches to Database Translation, NBS Special Publication 500-115, May 1984.

[9] Frank, W.L., Critical Issues In Software: A Guide to Software Economics, Strategy, and Profitability, John Wiley and Sons, New York, 1983.

[10] A Simple Guide to Five Normal Forms in Relational Database Theory, Communications of the ACM, Vol 26, No 2, February, 1983.

RECOMMENDED READING MATERIAL

[1] Martin, J. and Oxman S., Building Expert Systems: A Tutorial, Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[2] Energy Department, Analysis of Benefits and Costs (ABC's Guideline): A Manager's Guide to Analysis of Benefits and Costs, Vol. 2, June 1988.

[3] International Organization for Standardization, ISO 2024, 1987-05-19, Information Processing Systems - Open Systems Interconnection- Specification of Abstract Syntax Notation, (ASN.1), 1987.

[4] International Organization for Standardization, ISO 2025, 1987-05-19, Information Processing Systems - Open Systems Interconnection- Specification of Basic Encoding Rules for Abstract Syntax Notation, (ASN.1), 1987.

[5] Mayne, A. and Wood, M.B., Introducing Relational Database, NCC Publications, Oxford Road, Manchester, M1 7ED, England, 1987.

[6] Brodie, M.L., and Mylopolus, J., On Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Technologies, Springer-Verlag, Reading, Massachusetts, 1968.